I hereby certify that, on the date shown below, this correspondence is being transmitted via the Patent Electronic Filing System (EFS) addressed to Examiner Dana H. Shin at the U.S. Patent and Trademark Office.

Date: August 19, 2008

/Brian C. Cholewa/
Brian C. Cholewa, Reg. No. 58,392

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Glauco P. Tocchini-Valentini, *et al.*
Serial No.:	10/821,777
Filed:	April 9, 2004
For:	METHOD OF RNA CLEAVAGE AND RECOMBINATION
Group Art Unit:	1635
Examiner:	Dana H. Shin
Attorney Ref.:	911076.90023
Confirmation No.:	1445

DECLARATION UNDER 37 C.F.R. § 1.132

Mail Stop AF
Commissioner For Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, Glauco P. Tocchini-Valentini, on oath say and declare that:

1. I am the same Glauco P. Tocchini-Valentini who is a named inventor of the above-identified patent application. Since 1998, I have been the Scientific Director of the A. Buzzati-Traverso Campus, Monterotondo (Italy, Rome). As the Director, I operate the following institutions: European Molecular Biology Laboratory (EMBL) outstation on mouse genetics; International Centre for Genetic Engineering and Biotechnology (ICGEB) outstation on mouse genetics; European Mouse Mutant Archive (EMMA); CNR Institute for Cell Biology. I have studied RNA since 1962 and was a post-doctoral fellow with E. Peter Geiduschek, Francis Crick and Sydney Brenner. I have been a Professor of Molecular Genetics both at the University of Chicago and the University of Rome. A copy of my *Curriculum Vitae* is attached as Exhibit A.

2. I reviewed the March 20, 2008 Office Action issued by the US Patent and Trademark Office. I understand that Claims 1 and 4-17 are rejected as obvious over Fabbri S, *et al.*, "Conservation of substrate recognition mechanisms by tRNA splicing endonucleases," Science 280:284-286 (1998) in view of Santoro S & Joyce G, "A general purpose RNA-cleaving DNA

enzyme," Proc. Natl. Acad. Sci. USA 94:4262-4266 (1997). I submit this Declaration to provide evidence explaining that neither Fabbri *et al.* nor the art provided the knowledge or motivation to a skilled person that eukaryal tRNA endonucleases could cleave trans-formed BHBs in non-tRNA target RNA molecules.

3. Fabbri *et al.* lists eight (8) authors that are as follows: Stefania Fabbri, Paolo Fruscoloni, Emanuela Bufardeci, Elisa Di Nocola Negri, Maria Baldi, Domenica Gandini Attardi, Emilio Mattoccia, Glauco P. Tocchini-Valentini. I was the principal author of Fabbri *et al.* and drafted a majority of the manuscript.

4. In support of the obviousness rejection, the Examiner cited a paragraph on p. 285 of Fabbri *et al.*, which she indicated as providing knowledge and motivation to the skilled person that eukaryal tRNA endonucleases could cleave trans-formed BHBs in a non-tRNA target RNA molecule. Because I wrote that paragraph, I respectfully disagree.

5. At that time of Fabbri *et al.*, we were not at all interested in producing trans-formed BHBs. The paragraph on p. 285 simply referred to the exciting possibility that in nature there could exist different cis-formed BHB-like motifs that could function as substrates (other RNA substrates in my sentence) for some tRNA splicing endonucleases.

6. In addition, that paragraph addressed what is depicted in FIGS. 1-2 of Fabbri *et al.* FIG. 1 showed cleavage of pre-tRNAs having the following structures: (1) a 5'-terminal phosphate group; (2) an acceptor stem comprising a seven base pair stem made by the base pairing of the 5'-terminal nucleotides with the 3'-terminal nucleotides; (3) a CCA tail at the 3' end; (4) a D loop comprising a four base pair stem ending in a loop; (5) an anticodon loop comprising a five base pair stem whose loop contains the anticodon; and (6) a T loop comprising a five base pair stem. The pre-tRNAs were cis-formed and structurally distinct from those shown in the application by having the structures noted above.

7. FIG. 2 showed cis-formed mini-substrates, which are also distinct from those shown in the application by having a 6 nucleotide loop that is not present in the structures shown in the application). In RNA hairpins, like the cis-formed mini-substrates shown in Fabbri *et al.*, the sequence of the loop is essential for the formation and stability of the structure. The fact that cis-formed mini-BHBs were cleaved by eukaryal endonucleases does not immediately implicate that the skilled person could form an active trans-formed structure and predict that an eukaryal

endonuclease would subsequently cleave the trans-formed structure. In fact, it was not until 2001 that my group started to consider trans-formed structures and non-tRNA molecules as substrates for eukaryal endonucleases (*see, e.g.*, Fruscoloni P, *et al.*, EMBO Reports 2:217-221 (2001); attached as Exhibit B).

6. Furthermore, we found in subsequent years that even more relaxed forms of BHB are cleaved by enzymes like the one derived from *Sulfolobus sulfataricus* (*see*, Tocchini-Valentini G, *et al.*, PNAS, 102:8933-8938 (2005); Tocchini-Valentini G, *et al.*, PNAS, 102:15418-15422 (2005); and Tocchini-Valentini G, *et al.*, PNAS, 104:12300-12305 (2007); all attached as Exhibit C).

7. As for Santoro & Joyce, I find that the Examiner inadvertently mischaracterized what it teaches to the skilled person. Santoro & Joyce discussed catalytic nucleotides known as DNAzymes, which recognize substrates based upon Watson-Crick base pairing; they do not require protein enzymes. In contrast, tRNA endonucleases do not use Watson-Crick base pairing for cleavage. Because Santoro & Joyce relate to DNAzymes, it cannot remain relevant to the obviousness of the invention as a skilled person could not apply the teachings of Santoro & Joyce to Fabbri *et al.* and have achieved the invention.

8. According to 35 U.S.C. § 25 and 37 C.F.R. § 1.68, we hereby declare that all statements made herein of our own knowledge are true, and that all statements made on information and belief are believed to be true; and further, that these statements are made with knowledge that willful false statements, and the like so made, are punishable by fine or imprisonment, or both, under Section 1001, Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated this 19th day of August 2008

G P Tocchini-Valentini

Glauco P. Tocchini-Valentini

CURRICULUM VITAE

GLAUCO TOCCHINI-VALENTINI

Place of Birth: Piandimeleto (Pesaro), Italy

Date of Birth: 30 October 1937

Address: Via Cassia n.240, Rome, Italy

Married with 3 children.

University Studies Doctoral Degree in Pharmacy, University of Rome, 1959.
Doctoral Thesis: A Method for the Determination of Riboflavines

Post-doctoral Experience

1958-59 Institute of Biochemistry of the University of Rome (Italy) with Prof. A. Rossi-Fanelli

1960 EURATOM fellow at the "Institute für Strahlenbiologie", Karlsruhe (Germany) with Prof. K.G. Zimmer

1962-64 Post-Doctoral Fellow in the Department of Biophysics of the University of Chicago (USA) with Prof. E.P. Geiduschek and Prof. R. Haselkorn

1965-68 Post-Doctoral Fellow in the Division of Molecular Genetics, MRC Laboratory of Molecular Biology, Cambridge (UK) with S. Brenner and F.H.C. Crick

Periods Abroad

1975 Visiting Professor in the Department of Biology, UCSD, La Jolla, California (USA)

1980- Various stages from assistant to associate and then full professor (autumn quarter) in the Department of Molecular Genetics and Cell Biology of the University of Chicago, Illinois (USA)

Prizes and Awards

1996 XXIV Scanno Prize – Province of L'Aquila (Italy)

1997 "Gaetano Quagliariello" Plaque

1997 High Officer of the Order of Merit of the Italian Republic

1999 HFSP – 10th Anniversary Award

2002 Tartufari Prize – Accademia dei Lincei

Affiliations

- American Academy of Arts and Sciences – elected foreign member in 1994

- European Molecular Biology Organisation (EMBO) – elected member in 1969

- European Science and Technology Assembly (ESTA)– member 1994-98

- Human Frontier Science Program Organisation (HFSPO) - One of the founders of the Programme

- One of the founders of the Human Genome Organisation (HUGO). Member since 1990

- Academia Europaea – elected member in 2000

Professional Activities

In 1962, hired as a Researcher by the Italian National Research Council (CNR) and worked with Prof. Franco Graziosi

From 1968 to 1971, worked at the Institute of Genetics and Biophysics (LIGB) in Naples, Italy, directed by Prof. A. Buzzati-Traverso

In 1971, transferred to the Institute of Cell Biology in Rome, Italy, directed by Prof. Rita Levi-Montalcini. Since 1971, in charge of the Mechanisms of Gene Expression division of the Institute of Cell Biology.

Current Position

Since 9 May 1980, Director of the Institute of Cell Biology in Rome, Italy

Participation in Recent Conferences and Meetings

1996 EMBO Workshop "Frontiers of Molecular Biology" (Rome)

1998 Meeting "Comité Editorial des Compte-rendus de l'Académie des Sciences", (Paris)

1999 EU Consortium RNA Structure and Function, (Rome)

1999 The Mouse ENU Mutagenesis Workshop, Monterotondo (Italy)

2000 Seventh Session ICGEB – Board of Governors (New Delhi)

2000 A European Programme for Centres of Excellence in Life Science and Biomedicine (Brussels)

2000 European Science beyond 2000 (Lisbon)

2000 E-Biosci Meeting, Heidelberg (Germany)

2000 EU Meeting "Design of RNA domain, substrates inhibitors of tRNA-recognising proteins", Munster (France)

2000 "Molecular Genetics and Molecular Biology", Symposium in honour of Peter Geiduschek, Pacific Grove, California (USA)

2000 HFSP 23rd Meeting of the Board and Trustees (Strasbourg)

2000 HFSP 24th Meeting of the Board and Trustees (Strasbourg)

2000 Fifth Annual Meeting of the RNA Society, Madison, Wisconsin (USA)

2000 Europolis Workshop, "Scientific and Technological Institutions and the Knowledge-Based Economy in Europe" (Lisbon)

2000 International Colloquium on Life Sciences, Ethics, Economy and Society, Bordeaux (France)

2000 E-Biosci Governing Body Meeting, Grenoble (France)

2001 HFSP 25th Meeting of the Board and Trustees (Strasbourg)

2001 Mutagenesis, Cloning and Disease Models in the Mouse (Rome)

2001 HFSP Awardees Meeting, Turin (Italy)

2001 EMBC – Strategic Working Party and EMBO E-Biosci Committee (Heidelberg)

2001 CNR-UNESCO Annual Meeting - National Institute for Infectious Diseases "Lazzaro Spallanzani" (Rome)

2001 EMBO Building Grand Opening (Heidelberg)

2001 EU EMMA Press Conference (Rome)

2001 EMBL Finance Committee - E-Biosci Governing Body Meeting (Grenoble)

2001 EEC mice consortium Meeting, 16th August (Paris)

2001 EEC mice consortium meeting 23rd September (Paris)

2001 IAPSS Conference (Rome)

2001 EMMA works consortium meeting (Monaco)

2001 Workshop "The Biology of the Post-Genomic Era" Trieste, (Italy)

2001 ICGEB Board of Governors, Trieste (Italy)

2001 EMBC/EMBL Meetings (Heidelberg)

2001 HFSP 26th Meeting of the Board of Trustees (Strasburgo)

2001 Italian EMBO Members Meeting (Rome)

2002 E-BioSci Meeting (Frankfurt Airport)

2002 EMBC – Strategic Working Party (Heidelberg)

2002 EMMA Meeting, Monterotondo (Rome)

2002 EU-6 Framework – pre-meeting (Strasbourg)

2002 HFSP - 27th Meeting of the Board and Trustees (Strasbourg)

2002 EUMORPHIA- Consortium Meeting - Charles de Gaulle Airport (Parigi)

2002 40th Anniversary of EMBO (Heidelberg)

2002 EMBC - 4th Strategic Working Party (Heidelberg)

2002 HFSP - Second Annual Awardees Meeting, Ottawa (Canada)

2002 HFSP – 3rd Intergovernative Conference (Berlino)

2002 Oriel/E-Biosci Summer Meeting, Wieslock (Germany)

2002 EMBC-EMBL Meetings (Heidelberg)

2002 EC, EMMA, EBI - Meetings (Frankfurt Airport)

2002 EMMA Meeting (Munich)

2002 Extremophiles 2002, Naple (Italy)

2002 European Research Council 2002 (Copenhagen)

2002 EMBC – Strategic Working Party (Oslo)

2002 EMBL-EMBC Meetings, Hamburg (Germany)

2002 HFSP - 28th Meeting of the Board and Trustees (Strasbourg)

2002 E-BioSci/ORIEL Joint Meeting (Oxford)

2003 E-BioSci Electronic Information – Programme Committee Information (Heidelberg)

2003 EUMORPHIA Assembly (Strasbourg)

2003 IPPC-EMMA, Monterotondo (Rome)

2003 EMBL Strategic Forward Look (London)

2003 EMBC – Strategic Working Party (Paris)

2003 Life Sciences in the European Research Council (Paris)

2003 Future of Mouse Genomics in Europe (Frankfurt)

2003 EMMA Meeting, Monterotondo (Rome)

2003 HFSP 29TH Meeting of the Board of Trustees (Strasbourg)

2003 V European Symposium of the Protein Society (Florence)

2003 DNA: 50 years on the double helix (Cambridge)

2003 EMBC – Meeting of the Working Party on Special Project on Research Grants/Awards (Paris)

2003 EMBL Strategic Forward Look (Paris Airport)

2003 ELSF – Meeting on the European Research Council (Venice)

2003 34th Ordinary Session of the EMBC, EMBL and Finance Com. (Heidelberg)

2003 RNA 2003 – Eighth Annual Meeting of the RNA Society (Wien)

2003 New Frontiers in Nucleic Acid Research (Berlin)

2003 Mouse Functional Genomics in Europe (Bruxelles)

2003 ORIEL Partners – E-BioSci/Oriel Annual Workshop, Varenna (Italy)

2003 EMBL, Strategic Forward Look (Zurig)

2003 EMMA Meeting (Harwell, UK)

2003 EUMORPHIA – First Annual Meeting, Slough (UK)

2003 EMBO Members Workshop, Killarney (Ireland)

2003 Tenth Session of the Board of Governors of the ICGEB, Trieste (Italy)

2003 EMBL Council, EMBL Finance Comm., EMBC Meetings, Hinxton (UK)

2003 International Symposium RNA 2003, Kyoto (Japan)

2003 21st Anniversary of EMBO Journal (Heidelberg)

2003 HFSP Board of Trustees (Strasbourg)

2003 EUMORPHIA WP 10 (Rome)

2003 EUMORPHIA Management Meeting (Frankfurt Airport)

2004 EMBC Extraordinary Session (Heidelberg)

2004 EUMORPHIA WP Meetings (Rome)

2004 EUMORPHIA General Assembly (Munich)

2004 15th Anniversary of HFSP/HFSP Board of Trustess (Strasbourg)

2004 EMMA Meeting, Monterotondo (Rome)

2004 EMBC Strategic Working Party (Heidelberg)

2004 EUMORPHIA WP 9/10 (Rome)

2004 International Regulome Consortium (Ottawa)

2004 HFSP – Fourth Annual Awardees Meeting, Hakone (Japan)

2004 European Research Programme in Bioinformatics (Bruxelles)

2004 EMBO – 40th Anniversary Meeting (Heidelberg)

2004 EMBL Council and EMBC Meetings

2004 "In vivo and Transgenic Mouse Models" Conference (Berlin)

2004 EMBO Council (Heidelberg)

2004 EMBO – Sectorial Meeting Bioinformatics and Computational (Rome)

2004 Mouse Mutagenesis Mouse, Heathrow (UK)

2004 EUMORPHIA Annual Meeting, Heathrow (UK)

2004 E-BioSci/ORIEL Annual Workshop / E-BioSci Electronic Information Programme Committee Meeting, Hinxton (UK)

2004 EMBC 35th Ordinary Session, EMBL Council and Finance Comm. (Heidelberg)

2004 Mouse Resource Centers Roundtable, Bar Harbor, Maine (USA)

2004 ICGEB – Board of Governors (New Delhi)

2004 Working Group HFSP Journal – HFSP/BOT (Strasbourg)

2004 EUMORPHIA WP 9, 10, 14 (Rome)

2004 Conference: "Funding basic research in the life sciences: exploring opportunities for european synergies" (Bruxelles)

2005 Research Infrastructure in Life Sciences Meeting (Paris)

2005 EMMAinf 1st Meeting, Monterotondo (Rome)

2005 MUGEN Kick-off Meeting (Atene)

2005 HFSP Journal Working Group (Strasbourg)

2005 EUMORPHIA General Assembly (Madrid)

2005 PRIME SAG (Scientific Advisory Group) (Madrid)

2005 Focus Meeting on the use of animals in scientific research (Madrid)

2005 HFSP – Board of Trustess and HFSP Journal Working group (Strasbourg)

2005 EMBO Council (Strasbourg)

2005 EMBC Strategic Working Party
Workshop on EMBC/EMBO and Improving Life Science Research in Europe (Berlin)

2005 2nd Mouse Resource Centers' Roundtable (Rome)

2005 HFSP Fifth Awardees Annual Meeting - HFSP Journal Working Group – HFSP 34th Board of Trustees (Bethesda, Maryland USA)

2005 EUCOMM and PRIME Meeting (Frankfurt Airport)

2005 EMMAinf, 2nd Meeting (Stockholm)

2005 First EUCOMM Meeting (Frankfurt Airport)

2005 EMBC 36th Ordinary Session (Heidelberg)

2005 Priorities for Mouse Functional Genomics in Europe - Promotion and application of the EMPReSS phenotyping platform (Frankfurt Airport)

2005 MUGEN - 1st Technological Workshop on Transgenesis (Marsiglia)

2005 EMBC Strategic Working Party – EMBC Workshop (London)

2005 FIMRe (Federation of International Mouse Resourcers) Meeting (Washington)

2005 Eumorphia – EuroMouse Conference
PRIME 1st Open Scientific Meeting (Venezia)

2005 ICGEB Board of Governors (Trieste)

2005 9th International Symposium on Parkinson's Disease
Neuroscience 2005 – 35th Annual Meeting (Washington)

2005 EMBC-EMBL Meetings (Heidelberg)

2005 Eumorphia – WP 9 , 10 (Rome)

2005 ESFRI - Expert Group on Genomics et al.- 1st Meeting (Bruxelles)

2005 2nd EMMAinf Meeting – Monterotondo (Rome)

2005 Working Group for HFSP Journal (Strasbourg)
HFSP – 35th Board of Trustees (Strasbourg)

2005 EMBL/EMBO/CNR Mini-Symposium on "Genes and Behaviour" -
Monterotondo (Rome)

2006 ESFRI – Expert Group on Genomics et al. – 2nd Meeting (Bruxelles)

2006 EUMORPHIA – 3rd Annual Meeting (Barcellona)

2006 EUCOMM - Kick Off Meeting (Munich)

2006 ESFRI – Expert Group on Genomics et al. – 3rd Meeting (Bruxelles)

2006 EMBL Council (Heidelberg)

2006 HFSP 36th Board of Trustees
HFSP Publishing Board (Strasbourg)

2006 EMMAinf Meeting (Harwell)

2006 HFSP 37th Meeting of the Board of Trustees (Parigi)

2006 EMBL Council (Heidelberg)

2006 EuroPhenome Meeting: 1st and 2nd Phenotyping (Monaco)

2006 1st International Mugen Conference "Animal Models for Human
Immunological Disease" (Atene)

2006 EBI-EMMA Group Meeting (Bruxelles)

2006 Extraordinary EMBL Council and Finance Committee meetings (Heidelberg)

2006 EMBL – ATC , Groundbreaking Ceremony (Heidelberg)

2006 EUCOMM submeeting (Monaco)

2006 EMBC, EMBL Council and Finance Committee Meetings (Grenoble)

2006 ICGEB, Board of Governors (New Delhi)

2006 HFSP – 38th Board of Trustees (Strasbourg)

2006 5th EMMAinf Meeting – Monterotondo (Rome)

2007 EUCOMM Annual Review (Monaco)

2007 FP7 - First Meeting of the Programme Committee for the Theme 1 : Health (Bruxelles)

2007 EUMODIC – Start up Meeting (Barcellona)

2007 Workshop "The future of Research on Mouse Functional Genomics" (Bruxelles)

Scientific Societies, Commissions and National and International Working Groups

- European Molecular Biology Organisation (EMBO) – elected member in 1969

- EMBO Council – member since 1981

- EMBO Council - President, 1985-87

- EMBO Journal - Editor in Chief, 1985-87

- RNA, Editorial Board Member, 1995-96

- European Science and Technology Assembly (ESTA) - member 1994-98

- Human Frontier Science Program Organisation (HFSPO) - He is one of the founders of the Programme. HFSPO Council - Vice President, 1990-93

- Italian Society of Biophysics and Molecular Biology (SIBBM) - member

- Italian Society of Microbiology and Microbial Biotechnology (SIMGBM) - member

- Human Genome Organisation (HUGO) - member 1990-

- Human Genome Organisation Council (HUGO) - Council Member, 1991-93

- American Academy of Arts and Sciences - Elected Honorary Foreign Member, 1994

- Study Commission on the International Relations of the Italian National Research Council (CNR) -member

- Inter-ministerial Commission for the Co-ordination and Application of Biotechnology of the Italian Ministry of Health - member, 1995

- Scientific Committee of the "A. Buzzati-Traverso" Foundation – member, 1995

- Study Group for the Evaluation of the Activities of the Research Organs of the CNR – member, 1995

- Institute of the Italian Encyclopedia "G. Treccani" – Member of the Advisory Board "Frontiers of Life" series, 1995-

-Technical-Consulting Commission EUREKA Project, Italian Ministry for Universities and Scientific and Technological Research (MURST) – member, 1995-96

- International Association for the Promotion of Co-operation with Scientists from the Independent States of the Former Soviet Union (INTAS) – expert, 1996

- "Comptes rendus de l'Académie des Sciences" – member of the Editorial Committee – 1997-

- CNR Study Commission on Cloning – member, 1997

- CNR Study Commission on Bioethics – member

- Human Frontiers Science Program (HFSP), Member of the Board of Trustees, 1996

- National Committee for Biosafety and Biotechology, Presidency of the Council of Ministers, Work Group for the Evaluation of Biological Risks - member

- Italian National Commission UNESCO – member, 1998-

- MURST Work Group Biotechnology Sector– member since 1999

- Secretary General, European, Molecular Biology Conference (EMBC), 2000

- Academia Europaea – Elected Member, 2000

- CNR Representative at the Italian National Commission UNESCO – 2000

- MURST Post-Genome National Commission – member since 2000

- MURST National Commission for Neuroscience – member since 2000

- CNR Member of the Coordination Committee, 2003

- ESFRI (European Strategy Forum on Research Infrastructures) Member of the expert group, 2005-2007

- Member of Italian National Expert Group, Health-Cooperation Programme Committee, European Union Framework Programme 7, 2007

- ESFRI (European Strategy Forum on Research Infrastructures) Italian Delegate of the Ministry of University and Research, 2007

ORGANISATIONAL AND MANAGERIAL EXPERIENCE

Since 1980, G. Tocchini-Valentini has been Director of the CNR's Institute of Cell Biology in Rome. Since the onset, the Institute has been divided into Divisions and Services. It currently employs 52 persons in research activities (37 with Doctoral Degrees and 15 technicians) and 7 persons in administrative and secretarial work.
Since 1998 he has co-ordinated the organisation of the new international scientific Campus "Adriano Buzzati-Traverso" at Monterotondo (Rome, Italy), which was created thanks to a CNR project that has involved also the most important European scientific organisations. The Monterotondo Campus was created by the CNR for the purpose of developing and internationalising Italian biological and biomedical research, and is named after Prof. A. Buzzati-Traverso, the scientist who, while working for the CNR, brought modern molecular biology to Italy. The CNR transferred the Institute of Cell Biology (IBC) from Rome to the Monterotondo Campus, where research teams work in close collaboration with the international institutions present on the Campus: the EMBL (European Molecular Biology Laboratory) and EMMA (European Mouse Mutant Archive).

In Italy, he has been the Advisor for Biology of two Ministers for Scientific and Technological Research and has covered numerous organisational functions within the CNR. He has been Director of both Subprojects and of "Finalised" Projects.

In Europe, he collaborated in the development of framework programmes and has been a member of ESTA. In particular, he contributed to the development of the concept of infrastructures in Life Sciences. He has been a member of the EMBO Council, of the EMBL's Scientific Advisory Committee and is currently general secretary of the EMBC (European Molecular Biology Conference) and a member of the EMBL Council.
He is one of the founders of HUGO and has been a member of the organisation's Council.
He is one of the founders of HFSPO and has been a member of that organisation's Scientific Council. Currently he is a member of its Board of Trustees.
He has participated in several G7 meetings for Bioethics and was president of the "Human Genome" meeting.

He has taught at both undergraduate and graduate levels at the University of Rome and at the University of Chicago.

He has organised numerous courses and conferences on behalf of EMBO, UNIDO and HFSPO, and has given numerous seminars in Europe, the United States and Japan.

Positions of Responsibility in Foreign Research Projects (recent)

1996-1999	Holder and co-ordinator of the European Framework Programme 4 contract "EMMA-European Mouse Mutant"
1996-1999	Holder and co-ordinator of the European Framework Programme 4 contract "EMMA-European Mouse Mutant Archive"
1998-2000	Holder and co-ordinator of the European Framework Programme 4 contract "EMMA-Resource Database"
1998-2000	Co-holder of the European Framework Programme 4 contract "Design of RNA domains, Substrates or Inhibitors of tRNA Recognising proteins"
2001-2005	Co-holder of the European Framework Programme 5 contract "ORIEL – an Online Research Information Environment for the Life sciences»
2001-2004	Holder and co-ordinator of the European Framework Programme 5 contract "EMMAworks"
2001-2004	Holder and co-ordinator of the European Framework Programme 5 contract "EMMAnet
2002-2005	Co-holder of the European Framework Programme 5 contract "EUMORPHIA»
2004-2008	Holder and co-ordinator of the European Framework Programme 6 contract "EMMAinf"
2005-2009	Co-holder of the European Framework Programme 6 contract "MUGEN»
2005-2008	Co-holder of the European Framework Programme 6 contract "EURASNET»
2005-2008	Co-holder of the European Framework Programme 6 contract "EUCOMM»
2006-2008	Co-holder of the European Framework Programme 6 contract "EUMODIC»
2006-2008	Co-holder of the European Framework Programme 6 contract "CASIMIR»

Scientific Co-ordination Activities

1981-1985	Co-ordinator of Subproject 1 of the CNR's Finalised Project "Genetic Engineering and the Molecular Bases of Hereditary Diseases"
	Member of the CNR's National Consulting Committee on Biotechnology and Molecular Biology
1995	Study Commission of the Finalised Project "Applications of Molecular Genetics to Human Health"
1997	Co-ordinator assigned by Eurohorcs of the "Survey of the strengths and weaknesses of European science" for the biochemistry, molecular and cellular biology sectors
1997-	Member of the Scientific Committee of the President of the CNR
1999	Co-ordinator of the CNR-MURST Strategic Project "Molecular Genetics" (Law 449/97)
1998	Co-ordinator of the CNR Strategic Project "Basic Technologies of Post-genomics"
1999-	External member of the Scientific Council of the CNR Institute of Biochemical and Evolutionary Genetics at Pavia, Italy
2000-	Co-ordinator of the CNR-MURST Project "Functional Genomics" (Law 449/97)
2000-	Member of the Co-ordinating Unit of the MURST-CNR Sector Programme "Biomolecules for Human Health" (Law 95/95)
2001	Co-ordinator of the CNR Study Commission for the MURST Strategic Programme "Post Genome"
2002	Co-ordinator of the Italian Basic Research Fund (FIRB-MIUR) network project: "Mouse mutants, Phenotypes and Human Diseases - Mouse PHD"
2005	Co-ordinator of the Italian Basic Research Fund (FIRB-MIUR) International Project: "CNR-IBC/G. Armenise Harvard Medical School

DIDACTIC AND TRAINING ACTIVITIES

1977 Professor of General Genetics at the University of Rome "La Sapienza", Rome, Italy

1980- Professor of Developmental Genetics in the Department of Molecular Genetics and Cell Biology of the University of Chicago, Illinois, USA

National and International Courses: Organisation and Teaching

- EMBO Course on Genetic and Molecular Analysis of Development, Rome, October 8-23, 1979

- EMBO Workshop on Eukaryotic RNA Synthesis *in vitro*, Rome, October 1-3, 1981

- EMBO-NSF Workshop on Eukaryotic RNA Synthesis *in vitro*, Rome, September 25-27, 1983

- EMBO-NSF Workshop on Eukaryotic RNA Processing, Rome, May 22-26, 1985

- Workshop on RNA-Protein Interactions, Urbino, June 29-July 1, 1987

- Workshop on Molecular Genetics of Yeast, March 20-31,1989, ICGEB, Trieste

- Theoretical Course on Molecular Genetics of Yeast, April 9-13, 1990, ICGEB, Trieste

- HFSP Workshop on RNA-Protein Interactions, Urbino, September 22-27, 1991

- Theoretical Course on RNA Structure and Function, April 8-10, 1992, ICGEB, Trieste

- Theoretical Course on Yeast Molecular Genetics, April 12-15, 1992, ICGEB, Trieste

- Theoretical Course on Yeast Molecular Genetics, May 8-12, 1992, ICGEB, Trieste

- Theoretical Course on RNA Structure and Function, March 29 – April 1, 1993, ICGEB, Trieste

- Theoretical Course on RNA Structure and Function, March 28-31, 1994, ICGEB, Trieste

- EEC meeting on RNA in Biotechnology, January 1994, Roma

- Theoretical Course on RNA Structure and Function, May 2-5, 1995, ICGEB, Trieste

- Theoretical Course on RNA Structure and Function, April 1-5, 1996, ICGEB, Trieste

- "Istituto Pasteur-Cenci Bolognetti", Meeting on Structure and Function of RNA, April 18-19, 1997, Roma

- Theoretical Course on RNA Structure and Function, April 21-24, 1997, ICGEB, Trieste

- Theoretical Course on RNA Structure and Function, April 20-23, 1998, ICGEB, Trieste

- Theoretical Course on RNA Structure and Function, April 12-15, 1999, ICGEB, Trieste

- Second Mouse ENU Mutagenesis Workshop, October 1-3, 1999, Monterotondo

- Course on Cryopreservation of mouse germplasm, November 22-23, 1999, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 3-6, 2000, ICGEB, Trieste

- Symposium to Honour of Peter Geiduschek "Molecular Genetics and Molecular Biology", May 12-14, 2000, Asilomar Conference Center, Pacific Grove, California, USA

- Course on Cryopreservation of Mouse Germplasm, November 9-17, 2000, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 9-12, 2001, ICGEB, Trieste

- Workshop "The Biology of the Post-Genomic Era", October 15-17, 2001, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 11-16, 2001, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 8-11, 2002, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 17-22, 2002, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 7-10, 2003, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 9-14, 2003, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, March 29-April 1, 2004, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 14-19, 2004, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 4-7, 2005, ICGEB, Trieste

- Techniques and Approaches in Mouse Behavioural Phenotyping, April 12-14, 2005, (Strasbourg)

- EMBO Lecture Course (ELC)/Eumorphia Summer School (ESS) - "The Mouse as a model for human disease", September 17-23, 2005, Strasbourg

- Course on Cryopreservation of Mouse Germplasm, November 7 -10, 2005, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 10-13, 2006, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 13 -16, 2006, Monterotondo (Rome)

- Theoretical Course on RNA Structure and Function, April 2-5, 2007, ICGEB, Trieste

- Course on Cryopreservation of Mouse Germplasm, November 5 -8, 2007, Monterotondo (Rome)

PUBLICATIONS

G. Tocchini-Valentini is the author of 96 publications in international journals and of numerous chapters in books and conference proceedings

1. **Tocchini-Valentini,G.P.**: Cold Spring Harbor Symp. on Quant. Biol. 28: 203-204 (1963).

2. **Tocchini-Valentini,G.P.,** Stodolsky,M., Aurisicchio,A., Sarnat,M., Graziosi,F., Weiss,S.B. and Geiduscheck,E.P.: Proc. Natl. Acad. Sci. 50: 935-942 (1963).

3. Geiduscheck, E.P., **Tocchini-Valentini,G.P**. and Sarnat,M.T.: Proc. Natl. Acad. Sci. 52: 486-493 (1964).

4. Colvill,A.J., Kanner,L.C., **Tocchini-Valentini,G.P.**, Sarnat,M.T. and Geiduscheck,E.P.: Proc. Natl. Acad. Sci. 53: 1140-1147 (1965).

5. Martin,G.S. and **Tocchini-Valentini,G.P.**: Ed. Koningsberg and Bosch, Elsevier Publ. Cy. Amsterdam (1967).

6. **Tocchini-Valentini,G.P**. and Mattoccia,E.: Proc. Natl. Acad. Sci. 61: 146-151 (1968).

7. **Tocchini-Valentini,G.P.,** Marino,P. and Colvill,A.J.: Nature 220: 275-276 (1968).

8. Marino,P., Baldi,M.I. and **Tocchini-Valentini,G.P.**: Cold Spring Harbor Symp. on QUANT. BIOL. XXXIII: 125-127 (1968).

9. Di Mauro,E., Synder,L., Marino,P., Lamberti,A., Coppo,A. and **Tocchini-Valentini, G.P.**: Nature 222: 533-537 (1969).

10. Felicetti,L., **Tocchini-Valentini,G.P.** and Di Matteo,G.F.: Biochemistry, 8: 3428-3432 (1969).

11. **Tocchini-Valentini,G.P.**, Felicetti,L. and Rinaldi,G.M.: Cold Spring Harbor Symp. on Quant. Biol. XXXIV (1969).

12. Crippa,M. and **Tocchini-Valentini,G.P.**: Nature 226: 1243-1244 (1970).

13. **Tocchini-Valentini,G.P.**, Crippa,M.: Nature 228: 993-995 (1970).

14. **Tocchini-Valentini,G.P.**, and Crippa,M.: Cold Spring Harbor Symp. on Quant. Biol. 35: 737-742 (1970).

15. Arditti,R.R., Eron,L., Zubay,G. and **Tocchini-Valentini,G.P.**, Connaway,S. and Beckwith,J.: Cold Spring Harbor Symp. on Quant. Biol. XXXV: 437-442 (1970).

16. **Tocchini-Valentini,G.P.** and Crippa,M.: Second Lepetit Colloquium, North Holland, pp. 237-243, (1971).

17. Crippa,M. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. 68: 2769-2773 (1971).

18. Rinaldi,G., Felicetti,L. and **Tocchini-Valentini,G.P.**: Biochem. Biophys. Commun. 42: 441-444 (1971).

19. Bazzicalupo,P. and **Tocchini-Valentini,G.P.:** Proc. Natl. Acad. Sci. 69: 298-300 (1972).

20. Brown,R.D. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. 69: 1746-1748 (1972).

21. Brown,R.D., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Studia Biophysica 33: 91-102 (1972).

22. Brown,R.D., Mattoccia,E. and **Tocchini-Valentini,G.P.:** Federation of European Biochemical Societies, 8th Meeting, North Holland, 173-181, 1972.

23. **Tocchini-Valentini,G.P.**, Mahdavi,V., Brown,R.D. and Crippa,M.: Cold Spring Harbor Symp. on Quant. Biol. XXXVIII: 551-558 (1973).

24. Brown,R.D. and **Tocchini-Valentini,G.P.**: In: Methods of Enzymology, Ed. L. Grossman and K. Moldave, Academic Press, New York (1974).

25. Grinswold,M.D., Brown,R.D. and **Tocchini-Valentini,G.P.**: Biochem. Biophys. Res. Commun. 58: 1093-1103 (1974).

26. Tatò,F., Attardi Gandini,D. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. 71: 3706-3710 (1974).

27. Martini,G., Tatò,F., Gandini Attardi,D. and **Tocchini-Valentini,G.P.**: Bioch. Bioph. Res. Commun. 72: 875-879 (1976).

28. Mattoccia,E., Gandini Attardi,D. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. 73: 4551-4554 (1976).

29. Gandini Attardi,D., Martini,G., Mattoccia,E. and **Tocchini-Valentini,G.P.** Proc. Natl. Acad. Sci. 73: 554-558 (1976).

30. Baldi,M.I., Bazzicalupo,P. and **Tocchini-Valentini,G.P.:** Biochem. Biophys. Res. Commun. 73: 985-992 (1976).

31. Gandini Attardi,D., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Nature 270: 754-756 (1977).

32. Carrara,G., Gattoni,S., Mercanti,D. and **Tocchini-Valentini,G.P.**: Nucl. Ac. Res. 4: 2855-2870 (1977).

33. Baldi,M.I., Mattoccia,E., **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. 75: 4873-4876 (1978).

34. **Tocchini-Valentini,G.P.,** Gandini Attardi,D., Mattoccia,E., Baldi,M.I. and Carrara,G.: DNA synthesis: present and future. Molineux and Kohyama eds. pp. 1123-1143 (1978).

35. **Tocchini-Valentini,G.P.,** Baldi,M.I., Carrara,G., Gandini Attardi,D. and Mattoccia,E.: Cold Spring Harbor Symp. 43: 603-611 (1979).

36. Mattoccia,E., Baldi,M.I., Carrara,G., Fruscoloni,P., Benedetti,P. and **Tocchini-Valentini,G.P.**: Cell 18: 643-648 (1979).

37. Baldi,M.I., Benedetti,P., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Cell 20: 461-467 (1980).

38. Gandini Attardi,D., De Paolis,A. and **Tocchini-Valentini,G.P.**: J. Biol. Chem. 256: 3654-3661 (1981).

39. Otsuka,A., De Paolis,A. and **Tocchini-Valentini,G.P.**: Mol. Cell Biol. 1: 269-280 (1981).

40. Carrara,G., Di Segni,G., Otsuka,A. and **Tocchini-Valentini,G.P.**: Cell 27: 381-389 (1981).

41. Di Segni,G., Carrara,G. and **Tocchini-Valentini,G.P.**, Shoulders,C.C., Baralle,F.E.: Nucl. Ac. Res. 9: 6709-6722 (1981).

42. **Tocchini-Valentini,G.P.,** Mattoccia,E., Baldi,M.I., Ogden,R. and Pande,G.: Cold Spring Harbor Symp. on Quant. Biol. 885-892 (1982).

43. Hall,B.D., Clarkson,S.G. and **Tocchini-Valentini,G.P.**: Cell 29: 3-5, 1982.

44. Mattoccia,E., Baldi,M.I., Pande,G., Ogden,R. and **Tocchini-Valentini,G.P.**: Cell 32: 67-76 (1983).

45. Benedetti,P., Baldi,M.I., Mattoccia,E. and **Tocchini-Valentini,G.P.**: The EMBO Journal 2: 1303-1308 (1983).

46. Baldi,M.I., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Cell 35: 109-115 (1983).

47. Gandini Attardi,D., Margarit,I. and **Tocchini-Valentini,G.P.**: The EMBO Journal 4: 3289-3297 (1985).

48. Baldi,M.I., Mattoccia,E., Ciafré,S., Gandini,D. and **Tocchini-Valentini,G.P.**: Cell 47: 965-971 (1986).

49. Fabrizio,P., Coppo,A., Fruscoloni,P., Benedetti,P., Di Segni,G. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. USA 84: 8763-8767 (1987).

50. Geiduscheck,E.P. and **Tocchini-Valentini,G.P.**: Ann. Rev. Biochem. 57: 873-914 (1988).

51. Mattoccia,E., Baldi,M.I., Gandini Attardi,D., Ciafrè,S., and **Tocchini-Valentini,G.P.**: Cell 55: 731-738 (1988).

52. Carrara,G., Calandra,P., Fruscoloni,P., Doria,M. and **Tocchini-Valentini,G.P.**: Cell 58: 37-45 (1989).

53. Galli,G., Matteoni,R., Bianchi,E., Testa,L., Marazziti,D., Rossi,N. and **Tocchini-Valentini,G.P.**: Biochem. Biophys. Res. Commun. 173: 680-688 (1990).

54. Gandini Attardi,D., Baldi,M.I., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Methods in Enzymology 181: 510-517 (1990).

55. Cellini,A., Lacatena,R.M. and **Tocchini-Valentini,G.P.:** Nucl. Acids Res. 19: 997-1000 (1991).

56. Doria,M., Carrara,G., Calandra,P. and **Tocchini-Valentini,G.P.**: Nucl. Acids Res. 19: 2315-2320 (1991).

57. Baldi,M.I., Mattoccia,E., Bufardeci,E., Fabbri,S., and **Tocchini-Valentini,G.P.**: Science, 255: 1404-1408 (1992).

58. Oppi,C., Wagner,T., Crisari,A., Camerini,B. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. USA, 89:8268-8272, (Sept.) (1992).

59. Bufardeci,E., Fabbri,S., Baldi,M.I., Mattoccia,E. and **Tocchini-Valentini,G.P.**: The EMBO J., 12: 4697-4704 (1993).

60. **Tocchini-Valentini,G.P.,** Baldi,M.I., Gandini Attardi,D. and Mattoccia,E.: Gene, 135: 93-97 (1993).

61. Gros,F. and **Tocchini-Valentini,G.P.:** Nature 369: 11-12 (1994).

62. Lacatena,R.M., Cellini,A., Scavizzi,F. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. sci. USA 91: 10521-10525 (1994).

63. Carrara,G., Calandra,P., Fruscoloni,P. and **Tocchini-Valentini,G.P.**: Proc. Natl. Acad. Sci. USA 92: 2627-2631 (1995).

64. Fruscoloni,P., Zamboni,M., Panetta,G., De Paolis,A. and **Tocchini-Valentini,G.P.**: Nucleic Acids Research 23: 2914-2918 (1995).

65. Marazziti D., Golini E., Gallo A., Matteoni R. and **Tocchini-Valentini G.P.**: In: Human Genome Analysis Programme. (eds. Hallen and Klepsch, Ios Press, Amsterdam, pp. 28, (1995)

66. **Tocchini-Valentini, G.P.**: Nature (Book Reviews) 376: 563-564 (1995)

67. Wagner, T., Oppi, C. and **Tocchini-Valentini,G.P.**: FEBS Letters 365: 13-17 (1995).

68. Williams, K.P., Ciafrè,S. and **Tocchini-Valentini,G.P.**: The EMBO J. 14: 4551-4557 (1995).

69. Ficca, A.G., Testa, L. and **Tocchini-Valentini, G.P.**: FEBS Letters 377: 140-144 (1995).

70. Matteoni,R. and **Tocchini-Valentini, G.P.**: Acc. Naz. Lincei, pp.125-136 (1996).

71. Williams,K.P., Ciafrè,S. and **Tocchini-Valentini,G.P.**: In: Gli Oligonucleotidi Sintetici - Principi e Applicazioni. (A.Cascino, A.Nicolin eds., UTET Periodidi Scientifici) pp. 201-209 (1996).

72. Scarabino,D. and **Tocchini-Valentini,G.P.**: FEBS Letters 383: 185-190 (1996).

73. Di Nicola Negri,E., Fabbri,S., Bufardeci,E., Baldi,M.I., Gandini Attardi,D., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Cell 89: 859-866 (1997).

74. Marazziti,D., Golini,E., Gallo,A., Lombardi,M.S., Matteoni,R. and **Tocchini-Valentini,G.P.**: Genomics 45: 68-77 (1997).

75. Mannironi,C., Di Nardo,A., Fruscoloni,P. and **Tocchini-Valentini G.P.** Biochemistry 36: 9726-9734 (1997).

76. Medici,R., Bianchi,E., Di Segni,G. and **Tocchini-Valentini,G.P.**: The EMBO J. 16: 7241-7249 (1997).

77. Fabbri,S., Fruscoloni,P., Bufardeci,E., Di Nicola Negri,E., Baldi,M.I., Gandini Attardi,D., Mattoccia,E. and **Tocchini-Valentini,G.P.**: Science 280: 284-286 (1998).

78. Marazziti,D., Gallo,A., Golini,E., Matteoni,R. and **Tocchini-Valentini,G.P.**: Genomics 53: 315-324 (1998).

79. Mannironi C., Di Nardo A., Fruscoloni P. and **Tocchini-Valentini G.P.**: In: The many faces of RNA (Eggleston D.S. et al. eds), Academic Press Limited., London (1998).

80. Scarabino,D., Crisari,A., Lorenzini,S., Williams,K. and **Tocchini-Valentini,G.P.:** The EMBO Journal 18: 4571-4578 (1999).

81. Mannironi,C., Scerch,C., Fruscoloni,P. and **Tocchini-Valentini,G.P.**: RNA 6: 520-527 (2000).

82. Fruscoloni,P., Baldi M.I. and **Tocchini-Valentini,G.P.**: EMBO reports 2: 217-221 (2001).

83. Marazziti,D., Golini,E., Magrelli,A., Matteoni,R. and **Tocchini-Valentini, G.P.**: Current Genomics 2: 253-260 (2001).

84. Gandini Attardi D. and **Tocchini-Valentini G.P.** : RNA 8:904-912 (2002).

85. Fruscoloni P., Zamboni M., Baldi M.I. and **Tocchini-Valentini G.P.**: Proc. Natl. Acad. Sci. 100:1639-1644 (2003).

86. Deidda G., Rossi N. and **Tocchini-Valentini G.P.**: Nature Biotechnology 21:1499-1504 (2003).

87 Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, Berns A, Bradley A, Brown S, Carmeliet P, Chambon P, Cox R, Davidson D, Davies K, Duboule D, Forejt J, Granucci F, Hastie N, de Angelis MH, Jackson I, Kioussis D, Kollias G, Lathrop M, Lendahl U, Malumbres M, von Melchner H, Muller W, Partanen J, Ricciardi-Castagnoli P, Rigby P, Rosen B, Rosenthal N, Skarnes B, Stewart AF, Thornton J, **Tocchini-Valentini** G, Wagner E, Wahli W, Wurst W.: Nat Genet. Sep;36(9):925-7 (2004).

88. Marazziti D., Golini E., Mandillo S., Magrelli A., Witke W., Matteoni R. and **Tocchini-Valentini G.P.** –Proc. Natl. Acad. Sci. 101:10189-10194 (2004).

89. Brown SD, Chambon P, de Angelis MH; Eumorphia Consortium: Nat Genet, Nov;37(11):1155, 2005

90. Di Segni G., Borghese L., Sebastiani S. and **Tocchini-Valentini G.P.** – RNA 11:70-76 (2005).

91. Tocchini-Valentini G. D., Fruscoloni P. and **Tocchini-Valentini G.P**. - Proc. Natl. Acad. Sci. 102:8933-8938 (2005).

92. Tocchini-Valentini G. D., Fruscoloni P. and **Tocchini-Valentini G.P.** – Proc. Natl. Acad. Sci., 102:15418-15422 (2005).

93. FIMRe Board of Directors - FIMRe: Federation of International Mouse Resources: Mamm. Gen., 17:363-364 (2006).

94. Marazziti D., Mandillo S., Di Pietro C., Golini E., Matteoni R. and **Tocchini-Valentini G.P.**: Proc. Natl. Acad. Sci., 104: 9846-9851(2007)

95. Mattaj I.W. and **Tocchini-Valentini G.P.** –Nature, 447:377-378 (2007).

96. Tocchini-Valentini G.D., Fruscoloni P. and **Tocchini-Valentini G.P.** – Proc. Natl. Acad. Sci., 104: 12300-12305 (2007)

97. Aidinis V., Chandras C., Manoloukos M., Thanassopoulou A., Kranidioti K., Armaka M., Douni E., Kontoyiannis D.L., Zouberakis M., Kollias G. and the Mugen NoE consortium –Nucleic Acid Research, 36:D1048-D1054, 2008

98. Di Segni G., Gastaldi S. and **Tocchini-Valentini G.P.** –PNAS, 105: 6864-6869, 2008

Number of publications in the foremost scientific journals during his period as director of the Institute of Cell Biology (1980-2008)

Nature	3
Proc. Natl. Acad. Sci. USA	11
Cell	9
J. Biol. Chem.	1
Mol. Cell Biol.	1
Nucl. Acid Res.	5
EMBO Journal	6
Ann. Rev. Biochem.	1
Biochem. Biophys. Res. Comm.	1
Methods in Enzymol.	1
Gene	1
FEBS Letters	3
Genomics	2
Biochemistry	1
RNA	3
EMBO reports	1
Cold Spring Harbor Symp. Quant. Biol.	1
Current Genomics	1
Nature Biotechnology	1
Nature Genetics	2

EMBO reports

scientific report

Cleavage of non-tRNA substrates by eukaryal tRNA splicing endonucleases

Paolo Fruscoloni, Maria I. Baldi & Glauco P. Tocchini-Valentini*

Istituto di Biologia Cellulare, CNR, Campus 'A. Buzzati-Traverso', Via E. Ramarini, 32, 00016 Monterotondo Scalo (RM), Italy

Received October 25, 2000; revised December 11, 2000; accepted January 8, 2001

Eukaryal tRNA splicing endonucleases use the mature domains of pre-tRNAs as their primary recognition elements. However, they can also cleave in a mode that is independent of the mature domain, when substrates are able to form the bulge–helix–bulge structure (BHB), which is cleaved by archaeal tRNA endonucleases. We present evidence that the eukaryal enzymes cleave their substrates after forming a structure that resembles the BHB. Consequently, these enzymes can cleave substrates that lack the mature domain altogether. That raises the possibility that these enzymes could also cleave non-tRNA substrates that already have a BHB. As predicted, they can do so, both *in vitro* and *in vivo*.

INTRODUCTION

Accuracy in tRNA splicing is essential for the formation of functional tRNAs, and hence for cell viability. In both Archaea and Eukarya the specificity of splicing resides in recognition of tRNA precursors by tRNA splicing endonucleases (Belfort and Weiner, 1997; Trotta and Abelson, 1999). Archaeal tRNA splicing endonucleases cleave pre-tRNAs only using an RNA structure comprised of two bulges of three nucleotides each (where cleavage occurs) separated by four base pairs. This structure, called the bulge–helix–bulge (BHB) (Figure 1A, 2) (Daniels *et al.*, 1985; Diener and Moore, 1998), functions independently of the part of the molecules that constitutes the mature tRNA, so we refer to this type of recognition of the cleavage sites as being the mature-domain independent mode. In contrast, eukaryal tRNA splicing endonucleases require interaction with the mature tRNA domain for orientation, so we refer to that recognition as the mature-domain dependent mode (Mattoccia *et al.*, 1988; Reyes and Abelson, 1988).

Recognition of pre-tRNAs by eukaryal tRNA splicing endonucleases normally requires the mature tRNA domain, as well as a base-pair, called the anticodon–intron (A-I) pair (Figure 1A, 1), which is formed between nucleotides in the anticodon loop and the intron (Baldi *et al.*, 1992). The A-I pair must be at a fixed distance from the mature domain for cleavage to occur and cleavage near this base pair generates the 3′ splice site. An independent cleavage event, also at a fixed distance from the mature domain (usually at a purine), generates the 5′ splice site.

The two modes of substrate recognition are characterized by two distances. In the mature-domain independent mode the helix of the BHB sets the distance between the two bulges; in the mature-domain dependent mode the distance is fixed relative to reference in the mature domain.

While the subunit structures of the eukaryal and archaeal enzymes differ significantly (Trotta and Abelson, 1999), as do the superficial structures of the cleavage sites, we have demonstrated that both the *Xenopus* and yeast tRNA splicing endonucleases can operate in the mature-domain independent mode, characteristic of Archaea (Fabbri *et al.*, 1998). The results reported in this paper explain why the eukaryal endonucleases retain the ability to operate in the mature-domain independent mode when their natural substrates do not have a BHB.

RESULTS AND DISCUSSION

The artificial substrate pre-tRNA^{Archaeal} contains both a mature domain and a BHB (Figure 1A, 2). The eukaryal enzymes cleave the substrate with a two base-pair insert in the anticodon stem, 2bp∇as (Figure 1A, 4), only in the mature-domain independent mode (Fabbri *et al.*, 1998). The sites of cleavage by the eukaryal enzyme are fixed by recognition of local BHB structure rather than by reference to the mature domain.

The *Xenopus laevis* endonuclease can also cleave *in vivo* in the mature-domain independent mode. When pre-tRNA^{Archaeal} and pre-tRNA^{Archaeal} 2bp∇as were injected into *Xenopus* oocyte nuclei, both substrates were spliced and ligated. The size of the mature tRNA^{Archaeal} 2bp∇as, which is four bases longer than

*Corresponding author. Tel: +39 06 90600317; Fax: +39 06 90091261; E-mail: gtocchini@ibc.im.cnr.it

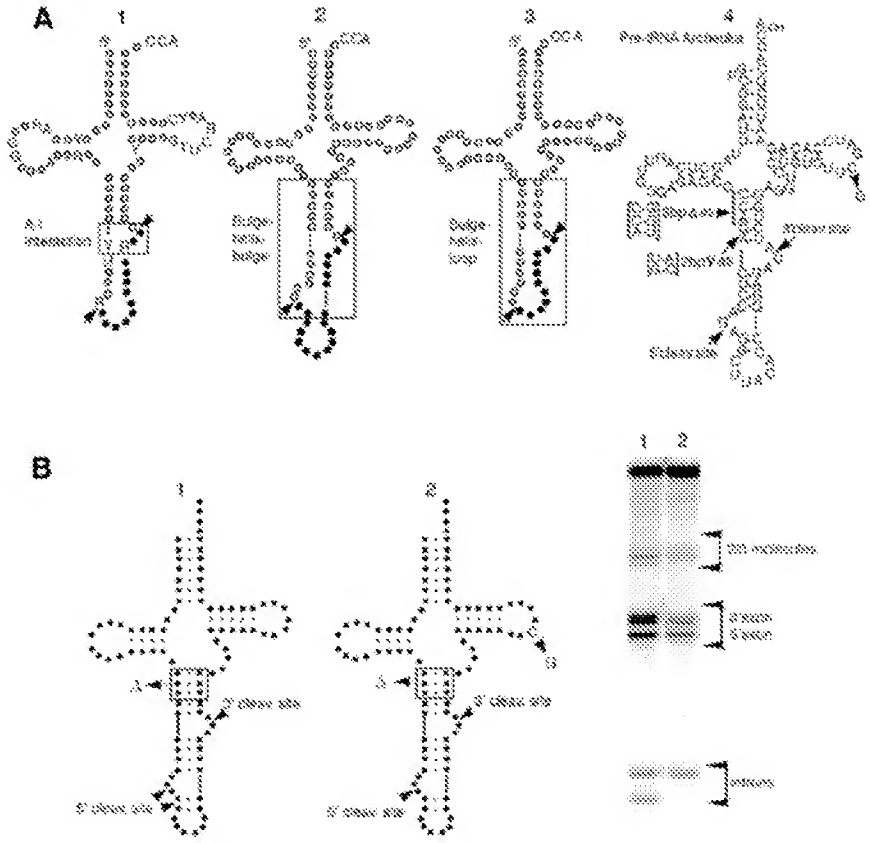

Fig. 1. (**A**) The A-I interaction. A conserved purine residue in the intron three nucleotides from the 3′ cleavage site (molecule 1, R in box) must pair with a pyrimidine in the anticodon loop 6 nucleotides upstream of the 5′ cleavage site (molecule 1, Y in box) to form the A-I (for anticodon–intron) interaction (Baldi *et al.*, 1992). BHB (molecule 2). Two bulges of three nucleotides each (where cleavage occurs) rigidly separated by four base pairs (Daniels *et al.*, 1985; Diener and Moore, 1998). BHL (molecule 3). A three-nucleotide 3′ site bulge, a four base-pair helix and a loop containing the 5′ site. Pre-tRNA^{Archaea} and its variants (molecule 4). The hybrid pre-tRNA molecule pre-tRNA^{Archaea} is a substrate for both the eukaryal and archaeal endonucleases. It consists of two regions derived from yeast pre-tRNA^{Phe} [nucleotides (nt) 1–31 and nt 38–76] joined by a 25 nt insert that corresponds to the BHB motif of the archaeal pre-tRNA^{Trp}. It has a typical eukaryal mature domain with cleavage sites located at the prescribed distance from the reference elements and a correctly-positioned A-I base pair, all of which should ensure correct recognition by the eukaryal endonuclease when the enzyme operates in the mature-domain dependent mode. In addition, the presence of the BHB motif confers substrate characteristics that are recognizable by the eukaryal enzyme when it operates in the mature-domain independent mode. (**B**) A substrate cleaved in both the mature-domain dependent and the mature domain independent modes. Products of digestion by the *Xenopus* tRNA splicing endonuclease. Molecule 1, pre-tRNA^{Archaea} 3bpΔas; molecule 2, pre-tRNA^{Archaea} 3bpΔas, C56G.

mature wild-type tRNA^{Phe}, indicates that the *Xenopus* enzyme cleaves in the mature-domain independent mode *in vivo* just as it does *in vitro* (Figure 2A). The substantial amounts of each precursor exported before cleavage probably results from saturation of nuclear retention (Arts *et al.*, 1998; Lund and Dahlberg, 1998).

Some substrates are cleaved in both modes. Pre-tRNA^{Archaea} 3bpΔas (Figure 1B), which has a three base-pair deletion in the anticodon stem, is cleaved in both modes. In this case, the two modes yield distinct product sizes, and both are observed. Two introns are visible in Figure 1B, lane 1. One of the products is not produced in the C56G mutant reflecting the inability of the enzyme to cleave a substrate that cannot form a normal mature domain in the mature-domain dependent mode.

We now propose that the orientation of the substrate in the active site of the eukaryal enzyme requires the formation of a structure that resembles a BHB; the A-I pair would play a pivotal role in this process, as it represents the closing base pair of one of the bulges. This model predicts that recognition of the mature tRNA domain by a eukaryal tRNA splicing endonuclease allows subsequent formation of a BHB-like cleavage structure.

In addition to the A-I pair (Baldi *et al.*, 1992), other relics of the archaeal world provide insight into the mechanism of the eukaryal cleavage reaction. Some eukaryal pre-tRNAs present motifs that resemble the BHB. The sequence of the *Caenorhabditis elegans* genome shows that tRNA genes corresponding to three isoacceptor species (Leu, Tyr, Ile) contain introns (The *C.elegans* Sequencing Consortium, 1998). The nematode pre-tRNAs present a motif which we call BHL (Figure 1A, 3), which resembles the BHB in that it has the 3′ site bulge and the four base-pair helix but the 5′ site is in a loop rather than in bulge. These three intron-containing pre-tRNAs of *C. elegans* are

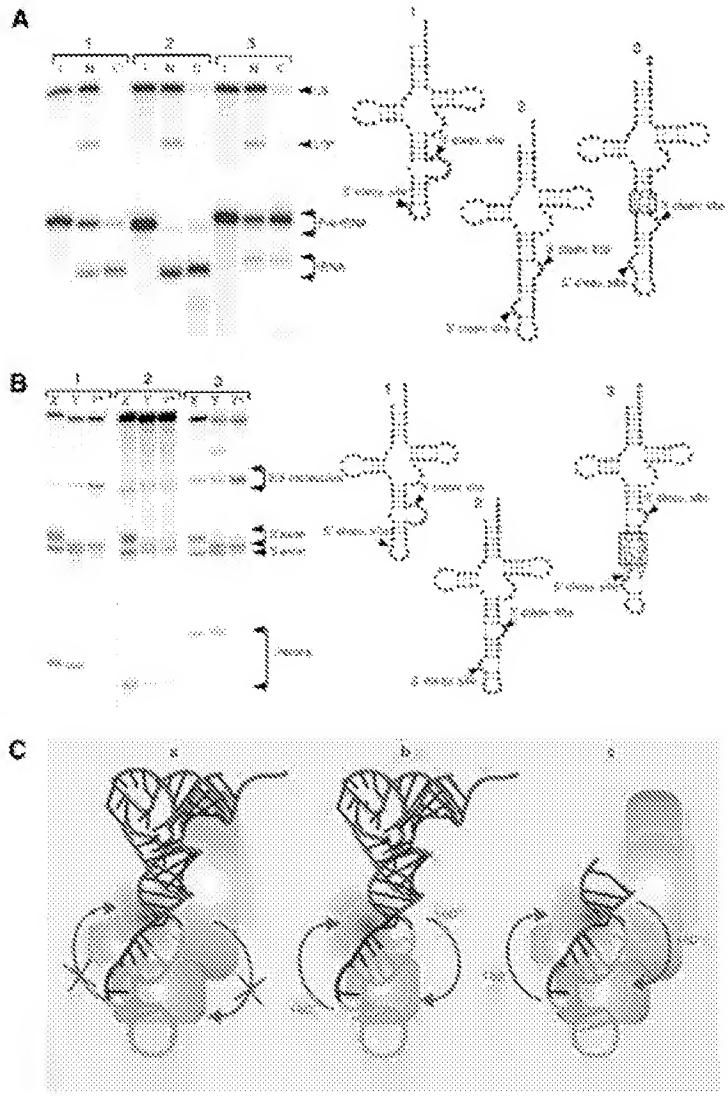

Fig. 2. (**A**) The *Xenopus* endonuclease can cleave *in vivo* in the mature-domain independent mode. Low amounts of ^{32}P-labeled RNAs corresponding to pre-tRNA^{Phe} (1); pre-tRNA^{Archeko} (2) and pre-tRNA^{Archeko}2bpV as (3) were injected into nuclei of *Xenopus* oocytes and 2 h later the intracellular distribution of the injected primary pre-tRNA transcripts and of the mature tRNAs were determined by analysis of total nuclear (N) and cytoplasmic (C) RNAs. The injected precursor RNAs in the cytoplasm probably resulted from inefficient nuclear retention. I, input. (**B**) The yeast endonuclease mutant sen2-3 cleaves pre-tRNA^{Archeko} at both sites. Molecule 1, pre-tRNA^{Phe}; molecule 2, pre-tRNA^{Archeko}; molecule 3, pre-tRNA^{Archeko} 3bpVBHB, C8G, G24C. X, *Xenopus laevis*; Y, yeast *S. cerevisiae* wild type; Y* yeast *S. cerevisiae* sen2-3. The sen2-3 preparation was contaminated with a 3′ exonucleolytic activity that partially degraded the 3′ end of the precursor, reducing the size of the 3′ half product. The sequences of the products of the intron excision reaction have been verified by fingerprinting (data not shown). In the sen2-3 mutant, one residue in loop L7, Gly^{92}, is changed to Glu (Trotta and Abelson, 1999). Loop 7 contains a histidine residue that is absolutely conserved in all tRNA endonucleases, and that probably acts as a general base by deprotonating the nucleophile 2′-hydroxyl group (Trotta and Abelson, 1999). The residues on loop 7 immediately surrounding the conserved histidine residue are not conserved among the tRNA endonucleases. We suggest that these residues have a role in the restructuring of the 5′ cleavage site in the eukaryal enzymes. (**C**) Comparison of models of enzyme-substrate interaction. (*a*) Pre-tRNA^{Archeko} Eukaryal enzyme (Trotta and Abelson, 1999). (*b*) Pre-tRNA^{Archeko} Archaeal enzyme (Trotta and Abelson, 1999). (*c*) BHB Eukaryal enzyme. A proposal for loss of symmetry during evolution of the intron excision reaction. In Archaea, the recognition element in pre-tRNA is the BHB, which has pseudo-two-fold symmetry (Diener and Moore, 1998; Trotta and Abelson, 1999). Since the endonuclease does not bind to the mature domain of pre-tRNA, the enzyme is oriented in such a way that both active sites can cleave either of the intron–exon junctions (*b*). The primary recognition element of the eukaryal endonuclease, on the other hand, is the asymmetrically located mature domain of pre-tRNA; interaction with that domain imposes an orientation of the enzyme on the substrate, so that each active site is specific to one or the other intron–exon junctions. In the absence of a mature domain, as with the mini-BHB (Fabbri *et al.*, 1998), the eukaryal enzyme is free to recognize pseudo-two-fold symmetric elements in the substrate, so that both active sites in the enzyme can bind to either junction (*c*). However, when a substrate has both a mature domain and a symmetric BHB, as in pre-tRNA^{Archeko}, the eukaryal endonuclease can interact with the mature domain, and the added energy of this binding would be likely to orient the enzyme (*a*). We propose that during evolution, once endonucleases were able to recognize the mature domain, the need for a symmetric BHB recognition site diminished; however, the active sites of the eukaryal enzymes were maintained, allowing them to cleave pre-formed BHB structures. Because the orientation of the two cleavage sites in the enzyme remained constant, the eukaryal pre-tRNAs had to maintain the ability to form a BHB-like structure upon binding the eukaryal enzyme; part of that requirement is seen in the A-I base pairing rule and in the BHL.

cleaved correctly by both yeast and *Xenopus* endonucleases, as well as *the Parascaris equorum* tRNA splicing endonuclease, but not by the archaeal enzyme (data not shown). Thus, the only truly universal substrate is an RNA with a BHB (Fabbri *et al.*, 1998).

Because they can both cleave the BHB, the archaeal and eukaryal endonucleases are likely to have identical dispositions of active sites, a feature conserved since their divergence from a common ancestor (Trotta and Abelson, 1999; Fabbri *et al.*, 1998) (Figure 2C). We suggest that the mature-domain dependent mode arose through specialization of the subunits of the eukaryal enzyme.

We propose that the eukaryal enzymes possess a mature-domain dependent 5′ site restructuring activity (Di Nicola *et al.*, 1997). Such an activity would be required for the last steps of substrate recognition by the eukaryal enzymes, recapitulating the recognition process of their archaeal counterparts. The 5′ site restructuring activity is not needed to cleave the BHB because it already has a correctly structured 5′ site; however, the activity is responsible for improving the efficiency of cleavage at the 5′ site in BHL (P. Fruscoloni, M. Zamboni, M.I. Baldi and G.P. Tocchini-Valentini, manuscript in preparation). The Ascaris enzyme also has a mature-domain dependent 5′ restructuring activity, but it differs from that of *Xenopus* because it is unable to restructure a typical eukaryal pre-tRNA such as yeast pre-tRNA^{Phe} (P. Fruscoloni, M. Zamboni, M.I. Baldi and G.P. Tocchini-Valentini manuscript in preparation).

Our model predicts the existence of mutants of the eukaryal enzyme that lack the 5′ restructuring activity. Such mutants would be unable to cleave a eukaryal pre-tRNA at the 5′ site, but could cleave at the 3′ site. More importantly, these restructuring mutants should cleave precursors that already have a BHB.

The yeast endonuclease is an αβγδ heterotetramer (Trotta *et al.*, 1997). Homology relationships and other evidence suggest that two subunits of the enzyme, Sen2p and Sen34p, contain distinct active sites, one for the 5′ site, the other for the 3′ site. The mutant sen2-3 is defective in cleavage of the 5′ site (Ho *et al.*, 1990); Figure 2B shows that sen2-3 extracts cleaves the 3′ but not the 5′ site of yeast pre-tRNA^{Phe}. The same extract, however, cleaves pre-tRNA^{Archaeka} at both sites (Figure 2B). Thus, sen2-3 cleaves the 3′ but not the 5′ sites in substrates lacking a BHB, as would be expected if it lacked the mature-domain dependent 5′ site restructuring activity. This conclusion is reinforced by the fact that pre-tRNA^{Archaeka} 3bp ΔBHB, a substrate that can interact with the enzyme only in the mature-domain dependent mode, is cleaved only at the 3′ site (Figure 2B).

It is unlikely that the lack of cleavage at the 5′ site of pre-tRNA^{Phe} results simply from inactivation of the catalytic site since the mutated amino acid is not near this site, based on the crystal structure of the enzyme (Li *et al.*, 1998). Moreover, pre-tRNA^{Archaeka}, which has a BHB, is cleaved at both sites even though its mature domain should prevent binding of the sen34 active site to the 5′ site (Figure 2B). Unfortunately, expression of a sen2-3 and sen34 double mutant enzyme is very likely to be lethal, making production of a doubly mutated enzyme impossible.

The ability of the eukaryal enzyme to recognize and cleave independently of the mature domain creates the possibility for cleavage of non-tRNA substrates (Figure 3). If the eukaryal endonuclease can recognize and cleave substrates in the mature-domain independent mode, any RNA that contains a BHB structure should be able to serve as a substrate. Such a target could be generated in mRNA by adding a suitable RNA oligonucleotide.

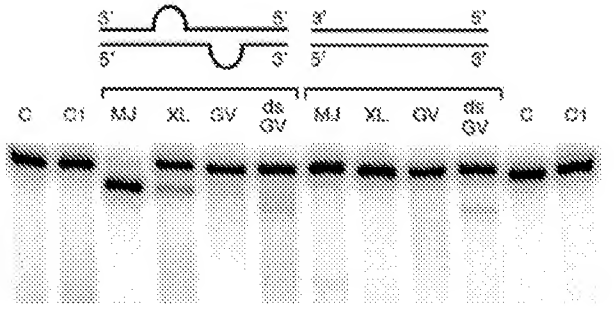

Fig. 3. Cleavage of a non-tRNA molecule by the *Xenopus* endonuclease. Profilin I mRNA duplexes (cartoon) consisting of ^{32}P-labeled sense strand and cold antisense strand (0.6 nM) were incubated with *Methanococcus jannaschii* endonuclease (MJ for 30 min at 65°C); *Xenopus laevis* endonuclease (XL for 90 min at 25°C); germinal vesicles extract (GV for 90 min at 25°C). The reacted RNA was treated as described (Mattoccia *et al.*, 1988; Baldi *et al.*, 1992; Fabbri *et al.*, 1998) and analyzed in 8M urea polyacrylamide gels. Two fragments are generated from profilin I mRNA (417 nts and 53 nts). The gel shows only the larger fragment. Unrelated ^{32}P-labeled dsRNA (low specific activity) was added where indicated (the concentration was 300× that of the profilin I duplex). C, duplex containing the BHB; C1, full duplex.

Figure 3 shows that the archaeal and eukaryal enzymes cleave mouse profilin I mRNA (Widada *et al.*, 1989), when the RNA is complexed with another oligoribonucleotide forming a BHB. This cleavage occurs in a BHB-dependent manner because fully double-stranded molecules (Figure 3) and molecules presenting an insertion of three base pairs in the helix of the BHB are not cleaved (data not shown). Figure 3 shows that cleavage also occurs in extracts of germinal vesicles (GV extracts). Again, cleavage is BHB-dependent. However, cleavage in this extract occurred only in the presence of a 100-fold excess of unrelated double stranded RNA (dsRNA). Pre-tRNA^{Archaeka}, on the contrary, is cleaved at high efficiency (data not shown). An explanation for these differences is the presence in GV extracts of adenosine deaminases (ADARs) that convert adenosines to inosines within dsRNA (Bass and Weintraub, 1988), thereby causing the RNA duplex to fall apart, disrupting the BHB structure. Presumably, at low concentration of dsRNA, ADARs deaminate the substrate and, as a result of the increased single-stranded character of the molecule, the BHB is destroyed.

Our results indicate that the formation of a BHB is an obligate step in cleavage by the eukaryal endonucleases and explain why the eukaryal endonucleases retain the ability to operate in the mature-domain independent mode when their natural substrates do not have a BHB.

METHODS

Pre-tRNAs were synthesized as described (Mattoccia *et al.*, 1988; Baldi *et al.*, 1992; Fabbri et al.,1998). Templates for the synthesis of the profilin I mRNA molecules were constructed by PCR. The PCR templates were the full-length molecules. One primer contained the T7 promoter. Conditions for PCR, transcription by T7 RNA polymerase, and endonuclease assays were

as described (Mattoccia *et al.*, 1988; Baldi *et al.*, 1992; Fabbri *et al.*, 1998). Duplex RNAs were prepared as described (Bass and Weintraub, 1988).

ACKNOWLEDGEMENTS

We are grateful to E. Mattoccia for continuous discussion and help with the figures. We thank J. Dahlberg for continuous encouragement and help with the manuscript; W. Witke for mouse profilin I cDNA; R. Haselkorn for critical reading of the manuscript; G. Deidda for help with the PCR and J. Abelson for the sen2-3 mutant. We thank A. Sebastiano for secretarial assistance, and G. Di Franco for technical assistance. Financial support was provided by: EC Contract BIO4-98-0189; Progetto Finalizzato 'Biotecnologie' (Consiglio Nazionale delle Ricerche); Programma Biotecnologie MURST-CNR Legge 95/95 5%; Programma Biomolecole per la Salute Umana MURST-CNR Legge 95/95 5%; Progetto Strategico 'Basi Fondamentali della Post Genomica' (Consiglio Nazionale delle Ricerche).

REFERENCES

Arts, G.-J., Kuersten, S., Romby, P., Ehresmann, B. and Mattaj, I.W. (1998) The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J.*, **17**, 7430–7441.

Baldi, M.I., Mattoccia, E., Bufardeci, E., Fabbri, S. and Tocchini-Valentini, G.P. (1992) Participation of the intron in the reaction catalyzed by the *Xenopus* tRNA splicing endonuclease. *Science*, **255**, 1404–1408.

Bass, B.L. and Weintraub, H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, **55**, 1089–1098.

Belfort, M. and Weiner, A. (1997) Another bridge between kingdoms: tRNA splicing in Archaea and Eukaryotes. *Cell*, **89**, 1003–1006.

Daniels, C.J., Gupta, R. and Doolittle, W.F. (1985) Transcription and excision of a large intron in the tRNA Trp gene of an archaebacterium, *Halobacterium volcanii*. *J. Biol. Chem.*, **260**, 3132–3134.

Di Nicola, E. *et al.* (1997) The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. *Cell*, **89**, 859–866.

Diener, J.L. and Moore, P.B. (1998) Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: The bulge-helix-bulge motif. *Mol. Cell*, **1**, 883–894.

Fabbri, S. *et al.* (1998) Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. *Science*, **280**, 284–286.

Ho, C.K., Rauhut, R., Vijayraghavan, U. and Abelson, J. (1990) Accumulation of pre-tRNA splicing '2/3' intermediates in a *Saccharomyces cerevisiae* mutant. *EMBO J.*, **9**, 1245–1252.

Li, H., Trotta, C.R. and Abelson, J.N. (1998) Crystal structure and evolution of a transfer RNA splicing enzyme. *Science*, **280**, 279–284.

Lund, E. and Dahlberg, J.E. (1998) Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science*, **282**, 2082–2085.

Mattoccia, E., Baldi, M.I., Gandini Attardi, D., Ciafrè, S. and Tocchini-Valentini, G.P. (1988) Site selection by the tRNA splicing endonuclease of *Xenopus laevis*. *Cell*, **55**, 731–738.

Reyes, V.M. and Abelson, J. (1988) Substrate recognition and splice site determination in yeast tRNA splicing. *Cell*, **55**, 719–730.

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**, 2012.

Trotta, C.R. and Abelson, J. (1999) tRNA splicing: An RNA world add-on or an ancient reaction? In Gesteland, R.F., Cech, T.M., Atkins, J.F. (eds), *The RNA World*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 561–583.

Trotta, C.R. *et al.* (1997) The yeast tRNA splicing endonuclease: A tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*, **89**, 849–858.

Widada, J.S., Ferraz, C. and Liautard, J.P. (1989) Total coding sequence of profilin cDNA from *Mus musculus* macrophage. *Nucleic Acids Res.*, **17**, 2855.

DOI: 10.1093/embo-reports/kve040

Structure, function, and evolution of the tRNA endonucleases of Archaea: An example of subfunctionalization

Giuseppe D. Tocchini-Valentini*†, Paolo Fruscoloni*, and Glauco P. Tocchini-Valentini*‡

*Istituto di Biologia Cellulare, Consiglio Nazionale delle Ricerche, Campus A. Buzzati-Traverso, Via Ramarini 32, I-00016 Monterotondo Scalo, Rome, Italy; and †Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche, Via Abbiategrasso 207, I-27100 Pavia, Italy

Edited by Robert Haselkorn, University of Chicago, Chicago, IL, and approved April 21, 2005 (received for review March 22, 2005)

We have detected two paralogs of the tRNA endonuclease gene of *Methanocaldococcus jannaschii* in the genome of the crenarchaeote *Sulfolobus solfataricus*. This finding has led to the discovery of a previously unrecognized oligomeric form of the enzyme. The two genes code for two different subunits, both of which are required for cleavage of the pre-tRNA substrate. Thus, there are now three forms of tRNA endonuclease in the Archaea: a homotetramer in some Euryarchaea, a homodimer in other Euryarchaea, and a heterotetramer in the Crenarchaea and the Nanoarchaea. The last-named enzyme, arising most likely by gene duplication and subsequent "subfunctionalization," requires the products of both genes to be active.

molecular evolution | RNA–protein interactions | tRNA splicing

Gene duplication is the primary source of new genes. The "subfunctionalization" hypothesis argues that duplicate genes experience degenerate mutations that divide the activity encoded by a single ancestral gene among its descendants (1). Here, we report a striking example of subfunctionalization.

In Archaea, the tRNA endonuclease plays a key role in assuring the correct removal of the intron from pre-tRNAs and pre-rRNA (2–6), which constitute the core of the translation machinery. Crystal structures of the tRNA endonucleases from *Methanocaldococcus jannaschii* (METJA) and *Archaeoglobus fulgidus* (ARCFU), both belonging to the phylum Euryarchaeota, are available (7, 8). These structures differ in a remarkable way. The structure of the homotetrameric endonuclease from METJA reveals two different functional roles for the monomeric units. The METJA endonuclease is organized as a dimer of dimers, with one subunit from each dimer participating in the catalytic cleavage reaction (the catalytic subunit) and the other (structural) subunit acting to place the two catalytic subunits correctly in space.

The crystal structure of the ARCFU endonuclease, by contrast, shows it to be a homodimer. Each subunit contains two similar repeating domains that are homologous to the subunit structure of the homotetrameric enzyme from METJA; the C-terminal repeat (CR) is the active domain, and the N-terminal repeat (NR) acts to stabilize the dimer.

The overall shape and size of the homodimeric ARCFU endonuclease resembles that of the homotetrameric METJA enzyme.

Both METJA and ARCFU belong to the Euryarchaeota. Nothing is known about the tRNA endonuclease of Crenarchaeota, the other main family of Archaea. To determine the properties of a crenarchaeal endonuclease, we searched the genome sequence of *Sulfolobus solfataricus* (SULSO) for homologs of the METJA endonuclease and found two candidate sequences.

Characterization of the two gene products reveals that both are required for tRNA endonuclease activity, each presumably functioning like one-half of the ARCFU enzyme. Detailed analysis of the amino acid sequences of the two proteins supports the idea that they evolved by the process called subfunctionalization (1, 9, 10).

Materials and Methods

Identification of the Homologs. The genes coding for the two SULSO proteins were PCR-amplified from genomic DNA by using two primers designed to obtain an amplified fragment presenting an NdeI site upstream of the gene and a BamHI site downstream. The digested PCR fragments were cloned in both pET28 (Novagen) and in pCYCA-11b (11). These two plasmids harbor two origins of replication that are compatible with one another and also two genes coding for different antibiotic resistances (kanamycin and chloramphenicol). The same procedure was used to clone the gene encoding the tRNA splicing endonuclease from ARCFU genomic DNA. The two truncated forms of endonuclease, NR_ARCFU and CR_ARCFU, were generated by PCR using the cloned full-length gene as template. NR_ARCFU comprises residues 1–138, and CR_ARCFU comprises residues 139–326. The PCR products were digested and cloned as previously described in pET28b and pCYAC-11b, respectively. All sequences of the clones obtained were verified.

Expression and Purification. Both proteins were overexpressed alone as hexahistidine-tagged protein (pET28) or coexpressed together with the untagged form (pCYCA-11b) in *Escherichia coli* BL21DE3 (Novagen). Cells were grown in a 1-liter culture of Luria–Bertani broth at 37°C in the presence of 30 μg/ml kanamycin (pET28) and adding, in the case of coexpression overproductions, 30 μg/ml chloramphenicol (pCYCA-11b). After cells reached an absorbance (at 600 nm) of 0.6, 1 mM IPTG was added to induce expression, and cells were grown for an additional 3 h. The bacterial pellets were resuspended in 50 mM Tris·HCl, pH 8/500 mM NaCl/1 mM PMSF/0.1% Tween 20 and lysed by adding lysozyme to a final concentration of 100 μg/ml. The extract was then heated at 65°C for 20 min and centrifuged for 1 h at 4°C. The supernatant was loaded on a 2-ml TALON cobalt affinity column (BD Biosciences) and washed with lysis buffer, and retained protein was eluted by adding 100 mM imidazole to the lysis buffer. We recovered large quantities of homogeneous protein "suitable for crystallization trials."

***In Vitro* Transcription.** The DNA template was prepared and transcribed by T7 RNA polymerase (12–14) using the T7-Megashort-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: METJA, *Methanocaldococcus jannaschii*; SULSO, *Sulfolobus solfataricus*; ARCFU, *Archaeoglobus fulgidus*; SULTO, *Sulfolobus tokodaii*; AERPE, *Aeropyrum pernix*; PYRAE, *Pyrobaculum aerophilum*; HALVO, *Haloarcula marismortui*; HALN1, *Halobacterium sp.*; METTH, *Methanothermobacter thermautotrophicus*; METKA, *Methanopyrus kandleri*; METAC, *Methanosarcina acetivorans*; METBA, *Methanosarcina barkeri*; METMA, *Methanosarcina mazei*; PYRAB, *Pyrococcus abyssi*; PYRFU, *Pyrococcus furiosus*; PYRHO, *Pyrococcus horikoshii*; FERAC, *Ferroplasma acidarmanus*; THEAC, *Thermoplasma acidophilum*; THEVO, *Thermoplasma volcanium*; NANEQ, *Nanoarcheum equitans*; CR, C-terminal repeat; NR, N-terminal repeat.

See Commentary on page 8791.

‡To whom correspondence should be addressed. E-mail: g.tocchini@ibc.cnr.it.

script kit (Ambion, Austin, TX). Transcripts were internally labeled with [^{32}P]UTP [800 Ci/mmol (1 Ci = 37 GBq), Amersham Pharmacia]. After phenol extraction and ethanol precipitation, transcripts were purified on a 10% denaturing polyacrylamide gel, eluted, and ethanol-precipitated.

***In Vitro* Splicing Assay.** Labeled tRNA precursors (20 fmol) were incubated with purified splicing endonucleases in reaction mixtures containing 25 mM Tris·HCl, pH 7.5/5 mM $MgCl_2$/100 mM NaCl/10% glycerol at 65°C for 1 h. Cleavage products were analyzed after phenol extraction and ethanol precipitation by gel electrophoresis on 10% denaturing polyacrylamide gel. Image analysis was done by using a Molecular Dynamics Model Storm 860 PhosphorImager with IMAGEQUANT 4 software.

Sequence Analysis. The gene coding for tRNA endonuclease subunits in METJA was used as the query sequence in BLASTP searches of the National Center for Biotechnology Information genomic database targeting the completed archaeal genomes including the species SULSO, *Sulfolobus tokodaii* (SULTO), *Aeropyrum pernix* (AERPE), *Pyrobaculum aerophilum* (PYRAE), ARCFU, *Haloarcula marismortui* (HALVO), *Halobacterium sp.* (HALN1), *Methanothermobacter thermautotrophicus* (METTH), METJA, *Methanopyrus kandleri* (METKA), *Methanosarcina acetivorans* (METAC), *Methanosarcina barkeri* (METBA), *Methanosarcina mazei* (METMA), *Pyrococcus abyssi* (PYRAB), *Pyrococcus furiosus* (PYRFU), *Pyrococcus horikoshii* (PYRHO), *Ferroplasma acidarmanus* (FERAC), *Thermoplasma acidophilum* (THEAC), *Thermoplasma volcanium* (THEVO), *and Nanonarcheum equitans* (NANEQ). The proteins detected were included in a multiple alignment built by using the program MAFFT (15). A refinement of the alignment was carried out by using the program RASCAL (16), and the quality was verified by NORMD (17).

Results

Genes encoding the tRNA endonucleases were sought in 20 completed archaeal genomes by using the METJA subunit sequence as the probe in BLASTP searches. We found a single gene that codes for a polypeptide of ≈180 residues in the genomes of PYRAB, PYRFU, PYRHO, METJA, and METTH. Presumably, the enzymes from these organisms assemble as homotetramers (α_4), like the METJA enzyme.

The single identified genes coding for an endonuclease in the genomes of METAC, METBA, METMA, ARCFU, HALVO, HALN1, FERAC, THEAC, and THEVO code for a subunit of ≈350 residues that presumably assembles into a homodimer, as does the ARCFU enzyme.

The search of the genomes of the Crenarchaeota SULSO, SULTO, AERPE, NANEQ, PYRAE, and METKA revealed two homologs. In the genome of SULSO, one gene codes for a protein, Q97ZY3, with an expectation value (*E* value) of $1e-22$, clearly indicating an ortholog. The size of this protein, 182 residues, suggests a possible assembly into a homotetrameric (α_4) enzyme, as in the case of METJA. We named this protein α_SULSO. The second homolog, Q980L2, has an *E* value slightly above the twilight zone of 0.001 (0.004). The protein encoded by this gene contains 177 residues. We reasoned that the second gene could code for a second subunit of the endonuclease. The METJA enzyme is a homotetramer, whereas the ARCFU enzyme is assembled as a homodimer with the monomeric unit resulting from duplication and fusion of a METJA-like monomer (18). We named the second SULSO protein β_SULSO. The SULSO enzyme might thus assemble as a heterotetramer: $\alpha_2\beta_2$.

Before describing biochemical experiments that explore the activities of the tRNA endonuclease subunits, we present a bioinformatic analysis of the amino acid sequences found and the structures predicted for these sequences.

We first characterized all of the sequences at the level of their primary structure (Fig. 1). The homodimer subunits were divided into two monomers of roughly the same size as the tetrameric monomer on the basis of the information obtained from the crystallographic structure of the ARCFU endonuclease. The NR (NR_ARCFU) is composed of residues 1–138, and CR (CR_ARCFU) is composed of residues 139–326. All of the protein sequences were then aligned as described in *Materials and Methods*.

The sequences were divided into two groups in relation to the role played in the enzymatic complex according to the ARCFU structure. One set included the structural subunits: the putative β-subunits and the NRs. The second set included the catalytic subunits: the α-subunit and the CRs. The homotetrameric subunits were included in both groups, because they can act as catalytic or structural subunits, depending on how they assemble. For clarity, the numbering of the residues used throughout refers to the METJA sequence. The multiple alignments of the structural and catalytic subunits are shown in Fig. 1. The similarities and identities (red and blue letters in Fig. 1) were separately determined for each group. Subsequently, homology regions were highlighted by taking into consideration the similarities and identities of the two distinct groups at the same time.

When the two sets are compared, three groups of residues can be distinguished: residues conserved between the two classes (yellow boxes), residues conserved only in the catalytic class (green boxes), and residues conserved in the structural class (cyan boxes). The majority of residues conserved in common between the two classes of subunits (yellow boxes) are clustered in two distinct segments of 15 aa: one segment includes strand β4-helix α2-strand β5 and the second includes helix α4-loop-strand β6. The first pattern presents a motif [L-x-L-x-x-(L,I,F,V)-x-Y-L-x-x-(K,R)-G-x-(L,I)] that is localized in the N-terminal domain, and the other presents a motif [(K,R)-(Y,F)-x_n-V-Y-x-D-L-(K,R)-x-(K,R)-G-(Y,F)-x-V] that is localized in the C-terminal domain. The first motif is not conserved in all of the structural subunits of crenoarcheote enzymes. In particular, in β-SULSO, this motif is highly divergent, and it has been deleted, along with the entire N-terminal domain, in β-SULTO and β-PYRAE. The second motif is universally conserved.

In the catalytic class (green boxes), the specific signature residues are clustered in loop L6, loop L7, and loop L9. The pattern of loop L6 is of six residues with a motif [(K,R)-(T,S)-(F,L,V,I)-K-(Y,F)-G]. Two of the three catalytic residues Y^{115} and H^{125} (residues in italics) are present in the conserved pattern of loop L7 [*Y*-x_n-*H*-(A,S)-(D,E)-(Y,F,W)-(I,L,V)], although the size of this loop is variable. L9 presents an eight-residue motif [R-(V,L)-(A,S)-H-(G,S)-(V,T)-R-*K*-(K,R)-(L,M)], which includes the catalytic residue K^{156} (in italics).

The structural subunit class (cyan boxes) is characterized by a [V-D-(E,D)-(E,D)-x-(D,E)-x-T] conserved pattern in loop L10, and a conserved E at position 135.

The mutually exclusive sequence motifs of the two classes can be represented on a surface calculated from the structures of the monomer (Fig. 2*A*) and of the homotetramer of the METJA endonuclease (Fig. 2*B*). Extension of the information available from the crystallographic studies to the endonuclease alignments strongly supports the existence of a canonical structure. This structure is formed by two different domains (Fig. 1). The N-terminal domain is characterized by a five-strand antiparallel β-sheet (strands β1, β2, β3, β4, and β5) packed against helix α2 and α3, both of which lie perpendicular to α1. The C-terminal domain is characterized by a five-strand β-sheet (strands β6, β7, β8, β9, and β10) sandwiched between helix α4 and α5.

The group of catalytic subunits shares a similar secondary structure (Fig. 1). The crenarchaeote proteins, in particular, contain an insertion in the loop between β3 and β4 rich in hydrophilic amino acids. CR_THEVO and CR_THEAC have lost helices α3 and α4. The size of loop L7 varies among all of the members of this

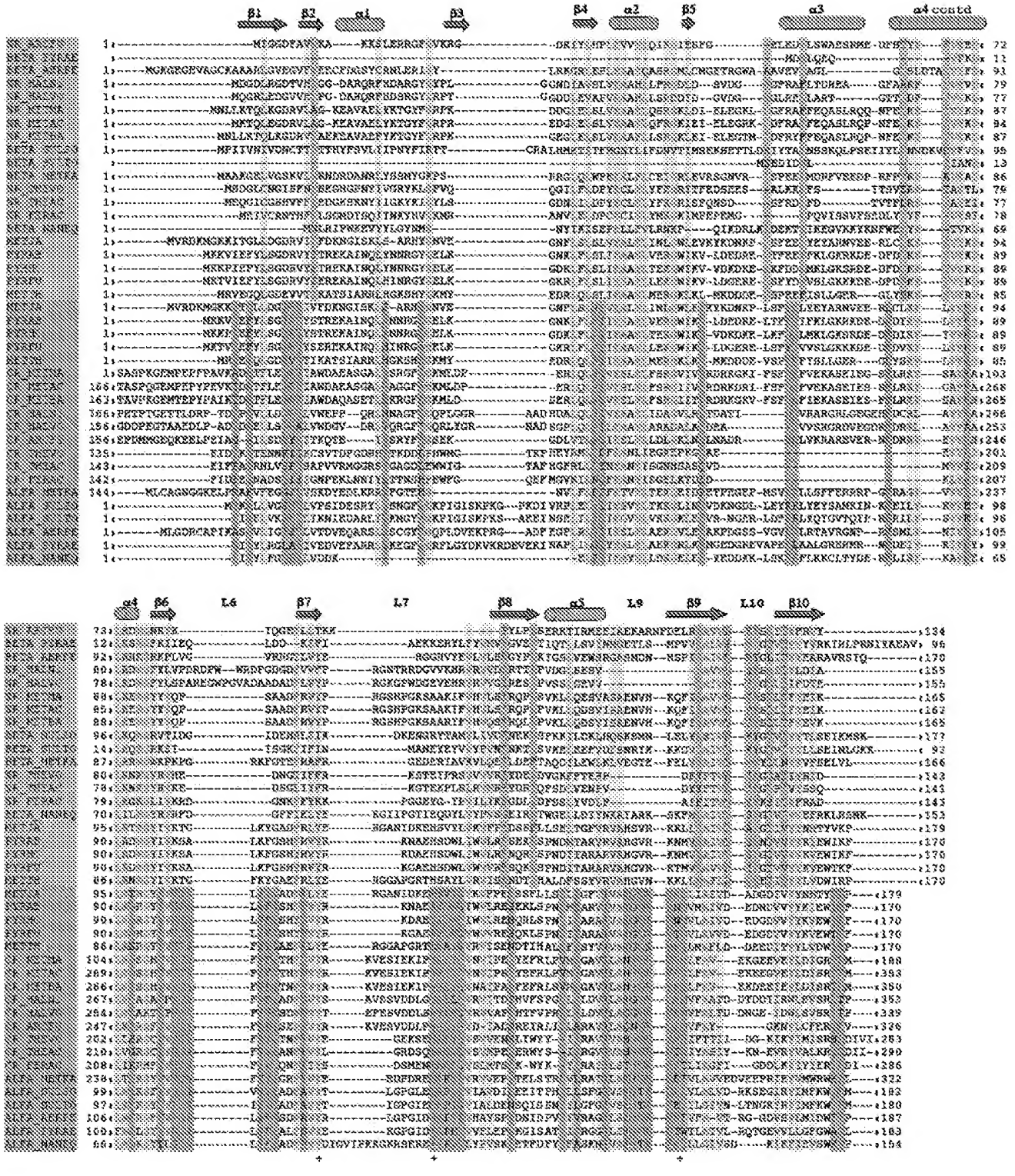

Fig. 1. Primary sequences and secondary structures of tRNA endonuclease subunits. A sequence alignment of tRNA endonuclease subunits is presented as two separate groups as described in the text. Sequences belonging to the structural subunit class are colored cyan, and sequences belonging to the catalytic subunit are colored green. Conserved residues are in red, and similar residues are in blue. Cyan columns indicate positions that are conserved among the catalytic subunits. Green columns highlight positions that are conserved among the structural subunits. Yellow columns indicate residues that are conserved in both groups. Secondary structure elements, as determined from the crystallographic structure of METJA, are indicated on top of the sequence of NR_ARCFU. Red arrows indicate β-strands, and green cylinders represent α-helices. The plus signs at the bottom of the alignment indicate the conserved catalytic triad Tyr-115/His-125/Lys-156.

group, but the two amino acids of the catalytic triads are always conserved. Variation in size and divergence in sequence are also present in loop L10.

The group of structural subunits shows, in the N-terminal domain, strong sequence divergence. In the case of β_SULTO and β_PYRAE, the entire N-terminal domain has been deleted.

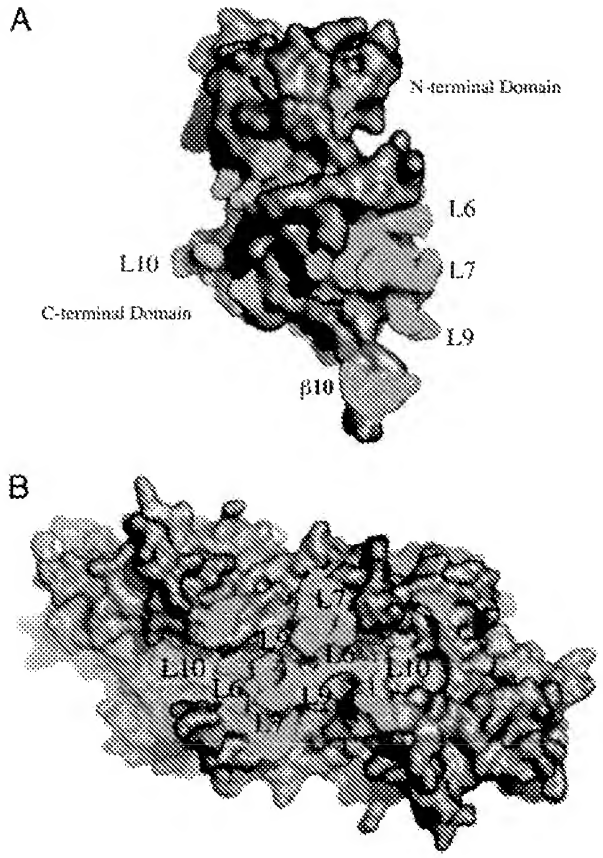

Fig. 2. Structural analysis. (*A*) Results of the multiple sequence alignment mapped onto the three-dimensional structure of a METJA monomer (1A79). Residues conserved in the structural subunits are colored cyan; residues conserved in the catalytic subunit are in green. (*B*) Sequence conservation mapped onto the structure of the tetramer of METJA (1A79).

NR_HALN1 and NR_HALVO show a larger loop between β6 and β7. The sequence and size of loop L7 is extremely varied in all of the class members. NR_ARCFU, in particular, shows a major deletion at this location. NR_THEAC, NR_THEVO, NR_FERAC, NR_HALVO, and NR_HALN1 are deleted in the loop between α5 and β9. It is evident from the alignment that loop L10 and the residues in the β10 strand, known to play a structural role in the dimerization of the ARCFU enzyme and the tetramerization of the METJA enzyme (7, 8), are conserved.

It is evident from this analysis and the published work of others that the known tRNA endonucleases of Archaea can take one of two forms: (*i*) a homotetramer in which two of the subunits play a structural role and two contain the catalytic sites or (*ii*) a homodimer, each of whose subunits, twice the length of the METJA class, contains a structural part and a catalytic part (7, 8, 18, 19). The bioinformatic analysis suggests that the latter enzyme arose from a gene duplication/fusion event, followed by a few mutations that allow for subfunctionalization of the two regions of the protein. To illustrate this possibility, we reversed the putative evolutionary path by cutting the gene encoding the ARCFU tRNA endonuclease into two segments, expressed each polypeptide independently, and assayed their ability to function *in vitro*.

For these experiments, we used the substrate tRNA archaeuka (Fig. 3*A*), a synthetic molecule containing an intron recognized by either archaeal or eukaryotic endonucleases (12, 20). Two truncated forms of the ARCFU enzyme, NR_ARCFU and CR_ARCFU, were generated. NR_ARCFU was cloned in pET28 to express a protein with an N-terminal His tag (His-6 NR_ARCFU). CR_ARCFU was cloned in pCYAC-11b. His-6 NR_ARCFU was overexpressed alone or in the presence of CR_ARCFU. After lysis, the crude extracts were pulled down by using a cobalt TALON affinity resin and extensively washed to remove nonspecifically bound proteins. The bound proteins were eluted from the resin with imidazole and assayed. Fig. 3*B* (lanes 4 and 5) shows that the excision of the intron from pre-tRNA archaeuka occurs only when the two truncated proteins are coexpressed. The purified proteins were visualized by Coomassie blue staining on SDS polyacrylamide gels (data not shown). These results are consistent with the conclusion that NR_ARCFU is the structural subunit, acting to place CR_ARCFU, the catalytic subunit, in space correctly.

Having shown that the two halves of a homodimer endonuclease could be assembled into an active enzyme, we investigated the endonuclease predicted from the genome sequence of the crenarchaeote SULSO. Recall that the genome analysis reveals one ORF (α) with very high similarity to the homotetramer of METJA and another with significantly lower similarity (β).

The gene for α_SULSO and the gene coding for the METJA monomer were cloned in pET28. The constructs code for proteins with a modified N terminus presenting a His-6 tag. The tagged enzymes were purified by affinity chromatography as described in *Materials and Methods*. The formation of aggregates was prevented by using high salt. Purification resulted, in both cases, in homogeneous proteins, as judged from Coomassie blue staining on SDS/PAGE. To determine whether the α_SULSO enzyme has a dual functional role like the METJA monomer, we tested the purified α_SULSO protein for cleavage of pre-tRNA archaeuka. We already knew that the intron of this substrate could be precisely excised by using a partially purified extract from SULSO (12). Fig. 3*C* (lane 6) shows that the substrate is correctly cleaved by the purified endonuclease from METJA but not by the protein α_SULSO (lane1). Leaving other possibilities aside for the moment, we conclude that α_SULSO cannot tetramerize like METJA. This observation led us to consider the possibility that β_SULSO allows the α-subunit to assemble properly.

The pull-down approach was then used to investigate the oligomerization of the two SULSO gene products. Two different constructs were produced for each gene to obtain an N-terminal-tagged version and an untagged version of the products. The tagged versions were overexpressed alone or in the presence of the untagged version of the other gene. After purification, the proteins were assayed by using pre-tRNA archaeuka as substrate. Fig. 3*C* (lanes 1 and 3) shows that neither β_SULSO nor α_SULSO alone cleaves the substrate; the activity is detectable only when the two proteins are coexpressed (Fig. 3*C*, lanes 2 and 4).

Thus, SULSO contains two genes, each encoding one subunit of a (presumably) heterotetrameric endonuclease, mimicking the specialization revealed in the assembly of the artificially created subunits of the ARCFU enzyme.

Discussion

A Previously Unrecognized Form of Endonuclease in SULSO. We present here the finding of a third archaeal tRNA endonuclease complex in the crenoarchaeote SULSO that is the result of assembly of two different subunits homologous to METJA, each one presumably playing one of the two different roles in a heterotetramer (Fig. 4). To identify the SULSO enzyme, we searched the sequenced genome by using METJA as a probe and found two homologs of the gene. The proteins encoded by these genes were expressed alone or together and, after purification to homogeneity, were tested for specific cleavage activity.

The enzyme is functionally active only if both subunits are present. Analysis of all of the completed genomes of 20 Archaea helped to extend the existence of a second homolog of the METJA gene to all Creanoarcheota and the Nanoarcheota and to deter-

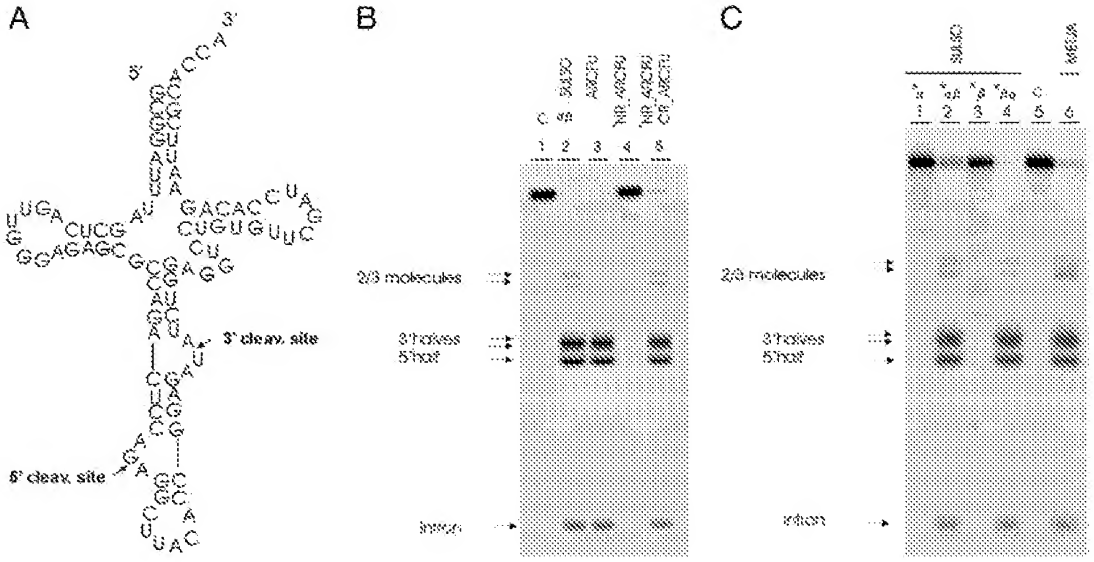

Fig. 3. tRNA endonuclease activity assays. (*A*) Pre-tRNA^{archaeuka} consists of two regions derived from yeast pre-tRNA^{Phe} (nucleotides 1–31 and 38–76) joined by a 25-nt insert that corresponds to the bulge–helix–bulge motif of archaeal tRNA^{Trp}. (*B*) Pre-tRNA archaeuka was incubated with four different recombinant proteins. The cleavage products were analyzed by electrophoresis on 10% polyacrylamide (29:1)/8 M urea gel, followed by autoradiography. The identification of the reaction products is indicated. Lane 1, control (no enzyme added); lane 2, endonuclease heterotetramer of SULSO; lane 3, endonuclease homodimer of ARCFU; lane 4, His-6-tagged NR_ARCFU domain alone; lane 5, NR_ARCFU coexpressed with CR_ARCFU domain. (*C*) Pre-tRNA archaeuka was incubated with five different recombinant purified proteins. Cleavage products were analyzed as in *B*. The identification of the reaction products is indicated. Lane 1, α_SULSO subunit; lane 2, α_SULSO coexpressed with the β_SULSO subunit; lane 3, β_SULSO subunit; lane 4, β_SULSO subunit coexpressed with the α_SULSO subunit; lane 5, control (no enzyme added); lane 6, METJA endonuclease. * indicates His-6 tagging.

mine the phylogenetic distribution of the different endonuclease architectures.

Evolution of tRNA Endonucleases in Archaea. The distribution of subunits in the different archaeal phyla and comparison of the aligned proteins provide a unique picture of how these enzymes evolved (Fig. 4). The ancestor of the archaeal enzyme was probably a homotetramer that, after two gene duplication events (or horizontal gene transfer), gave rise to a homodimeric form and a heterotetrameric form. The duplication events were independent from one another. One event took place in the common ancestor of Crenoarcheota and Nanoarcheota and resulted in two genes

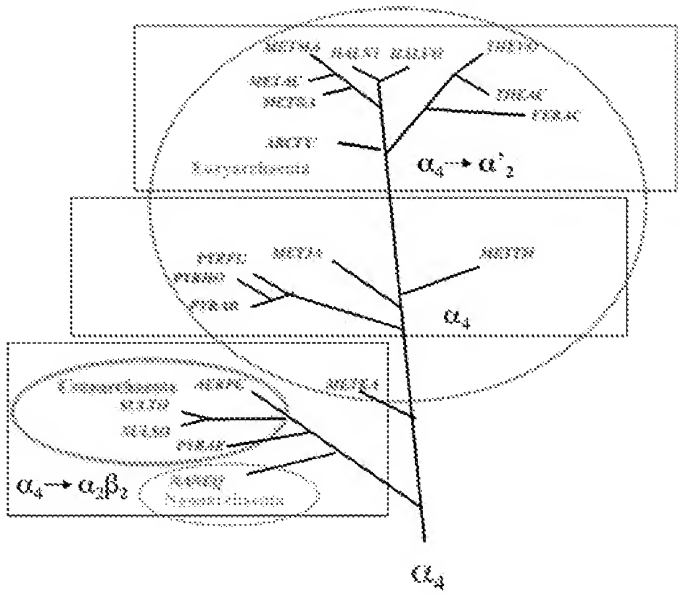

Fig. 4. Schematic representation of the phylogenetic tree of Archaea (adapted from ref. 21). Each rectangle surrounds species sharing the same endonuclease architecture. The subunits are named according to Fig. 5. The duplication events are indicated by an arrow.

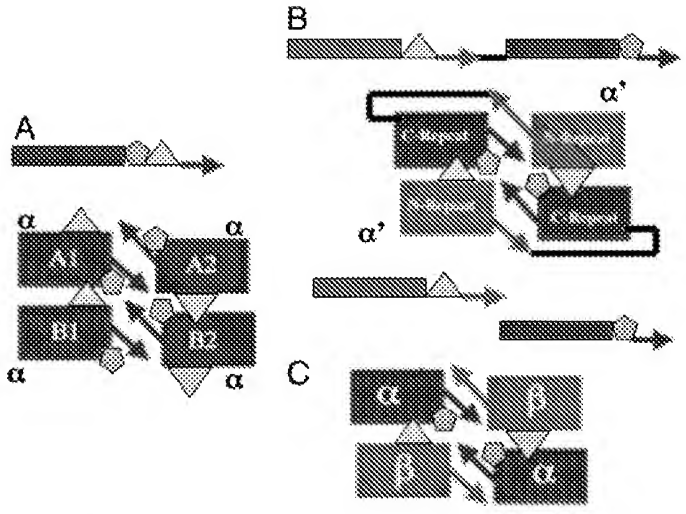

Fig. 5. Model of the tRNA splicing endonucleases of METJA, ARCFU, and SULSO. (*A*) model of the METJA homotetramer. Several important structural features discussed in the text are indicated: loop L10 (cyan triangle), the C-terminal β10 strand (arrow), and the conserved catalytic triad residues Tyr-115, His-125, and Lys-156 (green pentagon). (*B*) Proposed subunit arrangement of the ARCFU endonuclease. The NR lacks two of the three putative active-site residues. It does, however, contain many of the features of the C-terminal domain, which are important for structural arrangements of the enzyme: in particular, the L10 sequence. The C-terminal repeat contains all of the sequence features of the METJA enzyme. Black lines represent the polypeptide chain connecting the CRs and NRs. (*C*) A proposed model for the SULSO endonuclease. The α-subunit of the SULSO enzyme contains the conserved catalytic triad (green pentagon) and the C-terminal β10 strand (arrow). The β-subunit of the SULSO enzyme contains loop L10 (cyan triangle) and the C-terminal β10 strand (arrow).

coding for two subunits, whereas the other occurred in the ancestor of Archaeoglobales, Halobacteriales, and Methanosarcinales, resulting in an in-frame duplication and giving rise to a single gene coding for two fused subunits.

Gene Duplication in Archaeal tRNA Endonucleases: An Example of Subfunctionalization. Two pairs of subunits (A1 and A2, and B1 and B2) associate to form isologous dimers by means of extensive interactions between their β10 strands (Fig. 5*A*, arrow). The carboxyl half of β10 from one subunit forms main-chain hydrogen bonds with the symmetry-related residues of β10 from the other subunit (β10′), leading to a two-stranded β-sheet spanning the subunit boundary. The tetramer is formed by means of heterologous interactions between the two dimers. The main interaction between the two dimers is by means of the insertion of loop L10 (triangle in Fig. 5) from subunits A2 and B2 into a cleft in subunits B1 and A1 between the N- and C-terminal domains of each monomer. The interaction is primarily polar between acidic residues in loop L10 and basic residues in the cleft. This arrangement causes the two isologous dimers to be translated relative to each other by ≈20 Å, bringing subunits A1 and B1 much closer together than A2 and B2, which do not interact at all, and resulting in an arrangement of subunits in which only one symmetrically disposed pair of active sites can recognize the substrate. These weak interactions are likely to have been conserved in evolution because they lead to the required positioning of the two active sites.

The intersubunit β10–β10′ interaction observed in the METJA endonuclease is now an intrasubunit interaction in the ARCFU endonuclease dimer. ARCFU endonuclease also uses a similar L10 interaction, but to form a dimer (Fig. 5*B*). The three residues at the previously identified active site in the METJA endonuclease (His-125, Lys-156, and Tyr-115; pentagon in Fig. 5) are superimposable with those in the ARCFU endonuclease (His-257, Lys-287, and Tyr-246). This result is consistent with the fact that both enzymes act on the same bulge–helix–bulge motif in their pre-tRNA substrates. The connection between the NRs and the CRs is a fully extended 10-aa residue segment that wraps around the C-terminal domain of the CR before joining the N-terminal domain (Fig. 5*B*). The two CRs, C1 and C2, occupy the same spatial location as subunits A1 and B1 in the METJA endonuclease (in Fig. 5*B*). The two NRs, N1 and N2, occupy the same spatial location as the A2 and B2 subunits in the METJA endonuclease and thus play the same structural roles as the A2 and B2 subunits. Significantly, the NR has retained the L10 structure-mediating dimerization of the ARCFU endonuclease, equivalent to its role in tetramerization of the METJA endonuclease (Fig. 5*B*). The L10 structure is absent from the CR, suggesting that L10 has a role only in dimerization.

We generated two truncated forms of the ARCFU enzyme, NR_ARCFU and CR_ARCFU. The two halves of the homodimer endonucleases can be assembled into an active heterotetramer; NR_ARCFU is the catalytic subunit and contains the conserved catalytic triad (Fig. 5*B*). SULSO contains two genes, each coding for one of the subunits of a heterotetrameric endonuclease, mimicking the specialization revealed in the assembly of the artificially created subunits of the ARCFU enzyme. The α-subunit of the of the SULSO enzyme contains the catalytic triad, and the β-sunbunit contains the loop L10.

The general model of the process of duplication–degeneration–complementation proposes that degenerative mutations facilitate the preservation of the duplicated functional gene (9, 10). The gene, once duplicated, can follow three possible alternative evolutionary fates. After mutations in the regulatory regions or a null mutation in the coding region, the activity of one copy can be lost (nonfunctionalization). One copy can acquire mutations that result in the gain of a new function (neofunctionalization). Finally, both copies can accumulate mutations that lead to specialization (e.g., catalytic or structural), such that each copy is necessary to function as the single-copy ancestor gene (subfunctionalization).

Sequence analysis based on the available structural data has revealed two sets of mutually exclusive and substantially conserved motifs, allowing assignment of the endonuclease subunits to two classes according to their functional role. The homotetramer presents both sets of motifs, which correspond to a subunit capable of performing either function, according to how it is assembled in the enzyme. The homodimer and heterotetramer present domains or subunits subspecialized in one function or the other, presenting, therefore, one set of motifs or the other. The biochemical data we have presented here clearly support qualitative subfunctionalization in the case of the homodimers and heterotetramers. Both domains or subunits are required for the correct assembly of the enzyme to process the pre-tRNA precursor. This subfunctionalization is even more evident in the case of an ARCFU heterotetramer reconstituted by expressing the homodimer gene as two truncated forms. The resulting heterotetramer is active.

Conclusions

Our results suggest that on an evolutionary scale, the common ancestor of all of the archaeal enzymes was a homotetramer with four identical subunits that could play both the catalytic role and the structural role interchangeably, as in the contemporary methanogens. The appearance of a second copy of the gene resulted in subfunctionalization of the two copies, resulting in the splitting of the two roles, as in the case of homodimers and heterotetramers. The accumulation of mutations compromised one function or the other, as is evident from the conservation of signature motifs.

We thank Mose Rossi [Istituto di Biochimica delle Proteine-Consiglio Nazionale delle Ricerche (CNR), Naples] for providing the genomic DNA of SULSO, METJA, and ARCFU; A. Ferrara for secretarial assistance; and G. Di Franco for technical assistance. This work was supported by the Programma Biomolecole per la Salute Umana Ministero dell'Universitá e della Ricerca Scientifica e Tecnologica (MURST) CNR L. 95/95 5%; Fondo Investimenti Ricerca di Base Ministero Instruzione Universitá Ricerca (MIUR) of Italy; Progetto Genomica Funzionale L. 449/97 MIUR-CNR; Progetto Strategico Tecnologie di Base della Postgenomica CNR; Progetto Strategico Biotecnologie MURST-CNR; Progetto Strategico Genetica Molecolare L. 449/97 MURST; and European Networks of Excellence (EUMORPHIA and MUGEN).

1. Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. (2002) *Genome. Biol.* **3**, RESEARCH0008.
2. Kjems, J. & Garrett, R. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 439–443.
3. Thompson, L. D. & Daniels, C. J. (1988) *J. Biol. Chem.* **263**, 17951–17959.
4. Abelson, J., Trotta, C. R. & Li, H. (1998) *J. Biol. Chem.* **273**, 12685–12688.
5. Kjems, J. & Garrett, R. A. (1988) *Cell* **54**, 693–703.
6. Kjems, J., Jensen, J., Olesen, T. & Garrett, R. A. (1989) *Can. J. Microbiol.* **35**, 210–214.
7. Li, H., Trotta, C. R. & Abelson, J. (1998) *Science* **280**, 279–284.
8. Li, H. & Abelson, J. (2000) *J. Mol. Biol.* **302**, 639–648.
9. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. (1999) *Genetics* **151**, 1531–1545.
10. Lynch, M. & Conery, J. S. (2003) *J. Struct. Funct. Genomics* **3**, 35–44.
11. Fribourg, S., Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A. & Moras, D. (2001) *J. Mol. Biol.* **306**, 363–373.
12. Fabbri, S., Fruscoloni, P., Bufardeci, E., Di Nicola Negri, E., Baldi, M. I., Attardi, D. G., Mattoccia, E. & Tocchini-Valentini, G. P. (1998) *Science* **280**, 284–286.
13. Mattoccia, E., Baldi, I. M., Gandini-Attardi, D., Ciafre, S. & Tocchini-Valentini, G. P. (1988) *Cell* **55**, 731–738.
14. Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E. & Tocchini-Valentini, G. P. (1997) *Cell* **89**, 859–866.
15. Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005) *Nucleic Acids Res.* **33**, 511–518.
16. Thompson, J. D., Thierry, J. C. & Poch, O. (2003) *Bioinformatics* **19**, 1155–1161.
17. Thompson, J. D., Plewniak, F., Ripp, R., Thierry, J. C. & Poch, O. (2001) *J. Mol. Biol.* **314**, 937–951.
18. Kleman-Leyer, K., Armbruster, D. W. & Daniels, C. J. (1997) *Cell* **89**, 839–847.
19. Lykke-Andersen, J. & Garrett, R. A. (1997) *EMBO J.* **16**, 6290–6300.
20. Fruscoloni, P., Baldi, M. I. & Tocchini-Valentini, G. P. (2001) *EMBO Rep.* **2**, 217–221.
21. Woese, C. R. (1996) *Curr. Biol.* **6**, 1060–1063.

Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea

Giuseppe D. Tocchini-Valentini*†, Paolo Fruscoloni*, and Glauco P. Tocchini-Valentini*‡

*Istituto di Biologia Cellulare, Consiglio Nazionale delle Ricerche, Campus A, Buzzati-Traverso, Via Ramarini, 32, Monterotondo Scalo, 00016 Rome, Italy; and †Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche, Via Abbiategrasso 207, 27100 Pavia, Italy

Edited by Robert Haselkorn, University of Chicago, Chicago, IL, and approved September 8, 2005 (received for review August 5, 2005)

Members of the three kingdoms of life contain tRNA genes with introns. The introns in pre-tRNAs of Bacteria are self-splicing, whereas introns in archaeal and eukaryal pre-tRNAs are removed by splicing endonucleases. We have studied the structures of the endonucleases of Archaea and the architecture of the sites recognized in their pre-tRNA substrates. Three endonuclease structures are known in the Archaea: a homotetramer in some Euryarchaea, a homodimer in other Euryarchaea, and a heterotetramer in the Crenarchaeota. The homotetramer cleaves only the canonical bulge–helix–bulge structure in its substrates. Variants of the substrate structure, termed bulge–helix–loops, appear in the pre-tRNAs of the Crenarchaeota and Nanoarcheota. These variant structures can be cleaved only by the homodimer or heterotetramer forms of the endonucleases. Thus, the structures of the endonucleases and their substrates appear to have evolved together.

molecular evolution | RNA–protein interactions | splicing

In Archaea, the tRNA splicing endonuclease is responsible for the correct removal of introns from pre-tRNAs and is also involved in the processing of pre-rRNA and presumably certain pre-mRNA (1–4). An RNA motif consisting of a bulge–helix–bulge (BHB) is the universal substrate of the endonucleases from all archaeal lineages and eukaryotes (5). This motif has been shown by biochemical and NMR studies to be comprised of two bulges of three nucleotides symmetrically disposed on opposite strands and separated by a helix of four base pairs (6, 7). Although a consensus sequence has been derived (8), the conformation of this structure appears to be more relevant than its sequence (9).

The development of the genomics of Archaea made possible a characterization of the genes coding for pre-tRNA substrates (10) and the genes coding for the tRNA splicing endonucleases (11). Most introns of archaeal pre-tRNA genes are located in the anticodon loop, between nucleotides 37 and 38, the unique location of their eukaryotic counterparts. However, in several Archaea, mostly in Crenarchaeota, introns have been found at other positions: the anticodon stem and loop, the D- and T-loops, the V-arm, or the amino acid arm. Marck and Grosjean (10) renamed the BHB as hBHBh′, indicating with the new name that the canonical BHB motif should be enlarged to include two outer helices having at least two Watson–Crick base pairs. For introns located at 37/38 and elsewhere in the pre-tRNA, canonical hBHBh′ motifs were not always found. Instead, a relaxed hBH or HBh′ motif, including the constant central 4-bp helix H flanked by one helix (h or h′) with at least two Watson–Crick base pairs on either side, could be discerned (10).

We recently detected two paralogs of the tRNA endonuclease gene of *Methanocaldococcus jannaschii* (METJA) in the genome of the crenarchaeote *Sulfolobus solfataricus* (SULSO) (11). This finding led to the discovery of a previously unrecognized heterotetrameric form of the enzyme. The two genes code for two different subunits, both of which are required for cleavage of the pre-tRNA substrate. Thus, three different forms of tRNA endonuclease can now be recognized in the Archaea: a homotetramer in some Euryarchaea (such as METJA), a homodimer in other Euryarchaea (such as *Archeoglobus fulgidus*, ARCFU), and a heterotetramer in the Crenarchaeota (such as SULSO) and Nanoarchaeota. The heterotetrameric form of the enzyme, arising most likely by gene duplication and subsequent "subfunctionalization," requires the products of both genes to be active (12, 13).

Marck and Grosjean (10) were correct to recognize the several forms of the substrates but, missing the second subunit of the endonuclease from the Crenarcheota, they incorrectly assigned particular forms of the substrate to particular enzyme structures.

In the present article, we analyze the relationship of the intron-containing motif of the pre-tRNAs to the tRNA endonuclease architecture in the Archaea and show that the relaxed form of the substrate requires either the dimeric or the heterotetrameric endonuclease to be cleaved properly.

Materials and Methods

Expression and Purification of the Protein Constructs. The genes encoding the endonucleases from ARCFU, SULSO, and METJA were PCR-amplified from the respective genomic DNA using two primers designed to obtain an amplified fragment presenting an NdeI site upstream of the gene and a BamHI site downstream. The digested PCR fragments of the genes coding for ARCFU, METJA, and the α-subunit of SULSO were cloned in pET28b (Novagen), and the gene coding for the β-subunit of SULSO was cloned in pCYCA11b (14). All of the clones obtained were verified by sequencing. The proteins were overexpressed as hexa-histidine-tagged forms (pET28b) with the exception of β-SULSO (pCYAC-11b) that was untagged to be coexpressed together with the tagged form of α-SULSO (pET28b) in *Escherichia coli* BL21DE$_3$ (Novagen). Cells were grown in 1-liter cultures of Luria-Bertani broth at 37°C in the presence of 30 μg/ml kanamycin (pET28) with the addition of 30 μg/ml chloramphenicol (pCYCA-11b) in the case of coexpression. The proteins were purified on a metal affinity column as described (11). Homogeneity of the enzymes was assessed by Coomassie blue staining of SDS polyacrylamide gels. The tRNA endonuclease from the toad *Xenopus laevis* (XENLA) was purified according to ref. 15.

***In Vitro* Transcription and Splicing.** DNA templates prepared as described (11) were transcribed by T7 RNA polymerase by using the Ambion (Austin, TX) T7-Megashortscript kit in the presence of [α-^{32}P]UTP (800 Ci/mmol; Amersham Pharmacia). Products of the correct size were purified on a 10% denaturing polyacrylamide gel after phenol extraction and ethanol precipitation. Labeled tRNA precursors (20 fmol) were incubated with purified splicing endonucleases in reaction mixtures containing 25 mM Tris·HCl (pH 7.5), 5 mM $MgCl_2$, 100 mM NaCl, and 10%

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: BHB, bulge–helix–bulge; BHL, bulge–helix–loop; METJA, *Methanocaldococcus jannaschii*; SULSO, *Sulfolobus solfataricus*; ARCFU, *Archeoglobus fulgidus*; XENLA, *Xenopus laevis*.

‡To whom correspondence should be addressed. E-mail: gtocchini@ibc.cnr.it.

www.pnas.org/cgi/doi/10.1073/pnas.0506750102

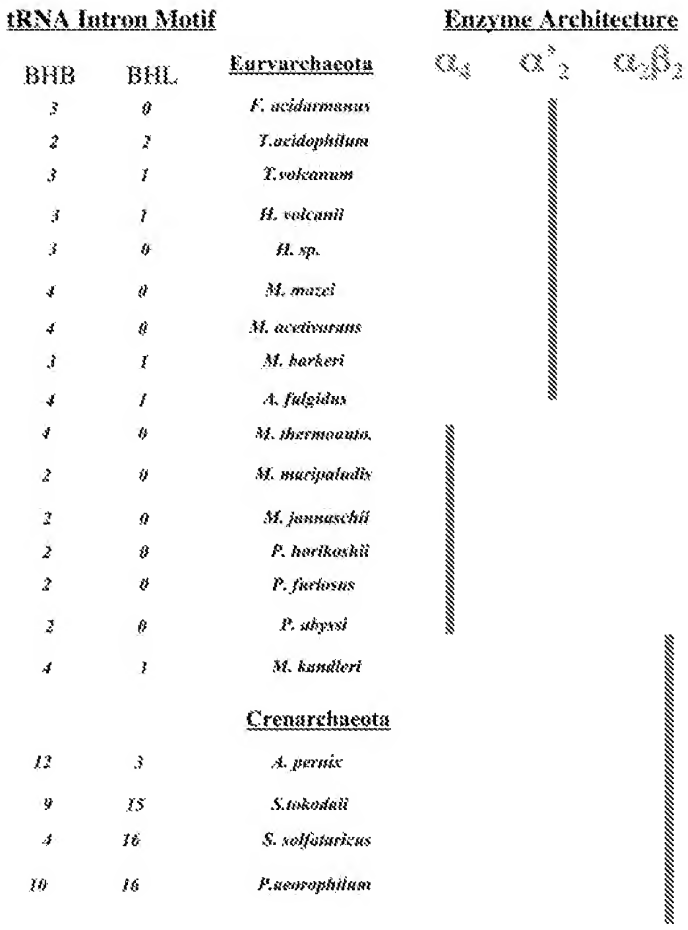

Fig. 1. tRNA intron motifs and enzyme architectures. Vertical bars indicate the species sharing the same endonuclease architecture (11). α_4 refers to the homotetramer, α_2 refers to the homodimer, and $\alpha_2\beta_2$ refers to the heterotetramer (11). The numbers indicate the number of hBHBh (BHB) and HBh or hBH (BHL) motifs present in the genome, according to ref. 10.

glycerol at 65°C for 1 h with the exception of the reaction containing the XENLA enzyme that was incubated at 22°C. Cleavage products were analyzed, after phenol extraction and ethanol precipitation, by electrophoresis on 10% denaturing polyacrylamide gels. Image analysis was done by using a Molecular Dynamics model Storm 860 PhosphorImager with IMAGE QUANT software, version 4.

Results

A Common Fold for All of the Archaeal Enzyme Subunits Is Stabilized by Two Conserved Residue Signatures. The genes coding for tRNA splicing endonucleases encoded in 19 different archaeal genomes have been described and characterized. Three different forms of the endonuclease have been distinguished; the ancestor of the archaeal enzyme was probably a homotetramer, which, after two independent gene duplication events (or horizontal gene transfer), gave rise to a homodimeric and a heterotetrameric form. One event took place in the ancestor of Crenarcheota, resulting in two genes coding for two subunits, whereas the other occurred in the common ancestor of Archaeoglobales, Halobacteriales, and Methanosarcinales, resulting in an in-frame duplication, giving rise to a single gene coding for two fused subunits (Fig. 1). We have shown how the homodimeric subunits can work as a heterotetramer, by cutting the gene into two independent segments each expressing a polypeptide (11). In all of the natural tetramers and the artificially generated one, each set of two subunits plays a specific role. One set contains the catalytic sites, and the other has the structural role of positioning the subunits with the active sites. These two different roles have resulted in the acquisition of mutually exclusive features that allow one to distinguish two functional classes of subunits (11).

Despite the existence of the two classes of subunits, a modular organization is conserved among them. The conserved residues in (β4)-(α2)-(β5) in the N-terminal domain and in (α4)-loop-(β6) in the C-terminal domain delineate motifs that provide a specific signature for the endonuclease family (Fig. 2). Using this universally conserved motif we can retrieve the subunits of Archaeal endonucleases selectively in the SwissProt database. The signature residues, represented in a cartoon model (Fig. 2*A*) of the METJA structure (16), stress the functional importance of helix α2 (blue) and helix α4 (purple) for the stabilization of each of the two domains and their positioning with respect to one another. When the residues are plotted every 100° consecutively around a spiral, the conserved residues in α2 are clustered on two opposite sides of the helix (Fig. 2*B*), presenting hydrophobic side chains. One face (residues 47, 50, 53, and 54) interacts with conserved residues in the β-sheet of the N-terminal domain. The other face (residues 48, 49, 52, and 56) interacts with the C-terminal domain and also directly with helix α4.

Helix α4 also presents conserved residues on two different faces (Fig. 2*C*). One face (residues 88, 91, 94, and 95) is packed against the β-sheet of the C-terminal domain, where it forms specific interactions with conserved residues. The other face (residues 85, 89, 92, and 96) interacts with helix α2 and the N-terminal domain. These observations support the existence of a canonical structure shared by all of the subunits, which implies that they all share a common origin (17, 18).

Canonical and Noncanonical Motifs in Intron-Containing Archaeal Pre-tRNAs. Following Marck and Grosjean (10), we examined the sequences spanning intron–exon junctions in intron-containing pre-tRNAs of 19 Archaea. Particularly interesting are those introns whose length is too short to form a second 3-nt bulge followed by a helix consisting of at least two Watson–Crick pairs. Fig. 3 shows that both hBH and HBh′ motifs are characterized by a bulge and an internal loop and can be represented by a structure that resembles the bulge–helix–loop (BHL), as described in some eukaryotic pre-tRNAs (19). Because, presumably, the archaeal endonucleases do not contact the mature domain, hBH and HBh′ do not appear as different to their enzymes. Hereafter, we shall refer to both hBH and HBh′ motifs as BHL-like motifs. Of 139 intron-containing pre-tRNAs, 82 contain a BHB and 57 contain a BHL–like motif. Fig. 1 shows that genes coding for intron-containing pre-tRNAs characterized by the BHL-like motif are absent from species that carry a homotetrameric (α_4) tRNA endonuclease. Pre-tRNAs containing BHL-like motifs are found only in those species characterized by the heterodimeric (α_2) or the heterotetrameric ($\alpha_2\beta_2$) forms of the tRNA endonuclease.

***In Vitro* Cleavage of Pre-tRNAs Presenting Either a BHB or a BHL Motif.** These observations lead to the prediction that α_4 endonucleases require a BHB substrate, whereas α_2 or $\alpha_2\beta_2$ endonucleases can cleave BHL substrates. These predictions were tested as follows: two different uniformly labeled pre-tRNA substrates were used for the cleavage assay (Figs. 4 and 5). These comprise a pre-tRNA presenting a motif with the intron and the boundary region of the 5′ exon and 3′ exon of the molecule folded into either one of a 2- or 3-nt bulge separated by a 4-bp helix (BHB motif) (Fig. 4*A*), and a pre-tRNA presenting a 3-nt bulge and an internal loop separated by a 4-bp helix (BHL-like) (Fig. 5*A*). We used as a control the partially purified endonuclease from XENLA, because it can process both substrates correctly, based on previous observations (19). Each substrate was incubated

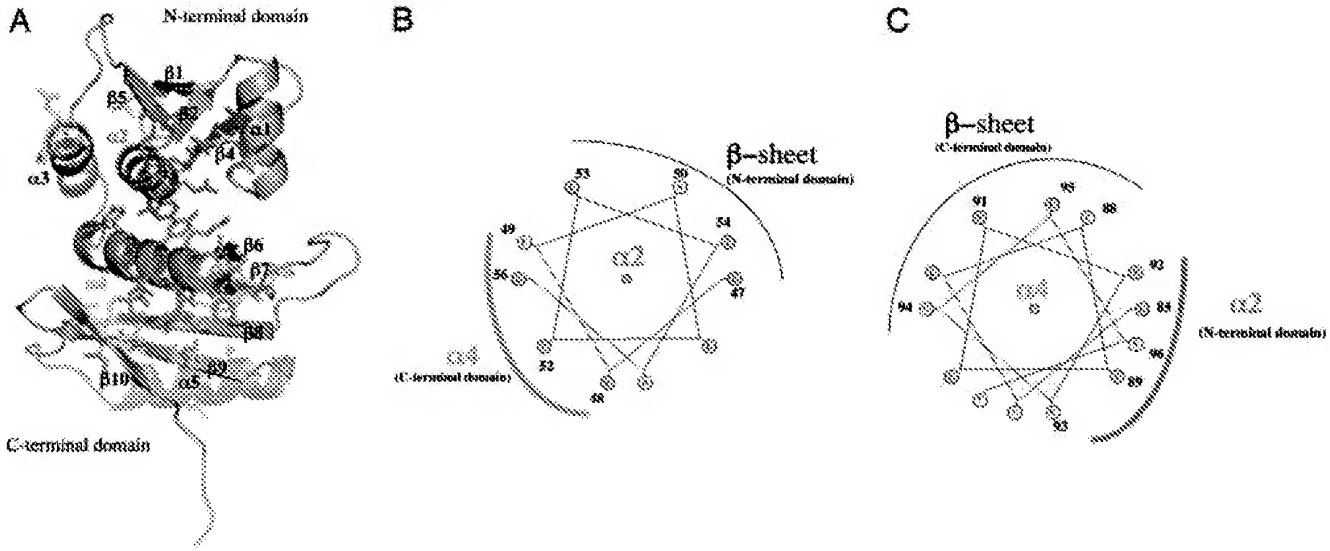

Fig. 2. Archaeal endonucleases common fold. (*A*) Cartoon representation of the monomeric subunit of METJA (16). All of the side chains represented are conserved among all of the subunits of archaeal endonucleases. The residues constituting the two signatures, as specified in the text, are colored blue and purple; the other conserved residues are yellow. (*B*) Helical wheel representation of helix α2. Hydrophobic residues are shaded in gray; only the conserved amino acids are numbered. (*C*) Helical wheel representation of helix α4. Representation as above.

with three different tRNA endonucleases from Archaea, representing each one of the three different architectures: the homodimer of ARCFU, the homotetramer of METJA, and the heterotetramer of SULSO.

Fig. 4*B* shows the expected result that all of the enzymes cleave the BHB substrate correctly. On the contrary, there are differences in the processing of the BHL-like intron. Fig. 5*B*, lane 2 clearly shows that the homotetrameric enzyme from METJA cannot cleave the BHL-like structure correctly. The enzyme cleaves inefficiently, only at the 3′ site, resulting in accumulation of the 5′ half-intron product, whereas the heterotetrameric enzyme from SULSO and the homodimeric enzyme from ARCFU cleave the BHL-like structure correctly (Fig. 5*B*, lanes 3 and 4), as does the (eukaryotic) XENLA endonuclease (Fig. 5*B*, lane 5).

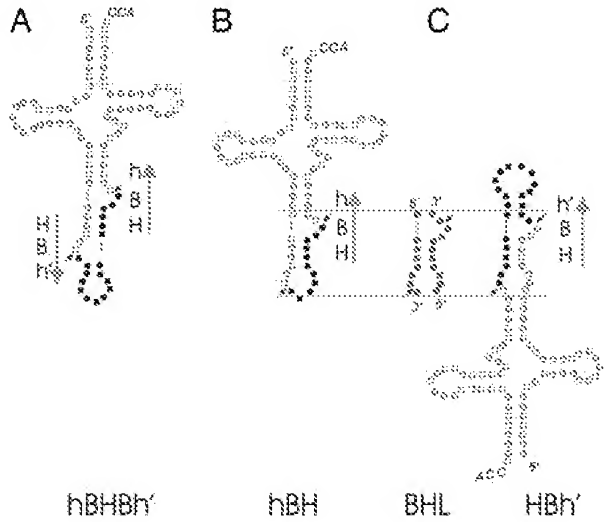

Fig. 3. Canonical and noncanonical motifs. (*A*) hBHBh′, BHB splicing motif flanked by two helices (h, h′) presenting at least two Watson–Crick base pairs. Arrows indicate the 5′ 32 → 3′ sense. (*B* and *C*) hBH and HBh′ relaxed BHB motifs. The two superimposed motifs are shaded in gray to show the similarity to a BHL motif.

Cis and Trans Splicing Motifs. All of the substrates described above were characterized by BHB or BHL targets produced by annealing strands belonging to the same RNA molecule. We turn

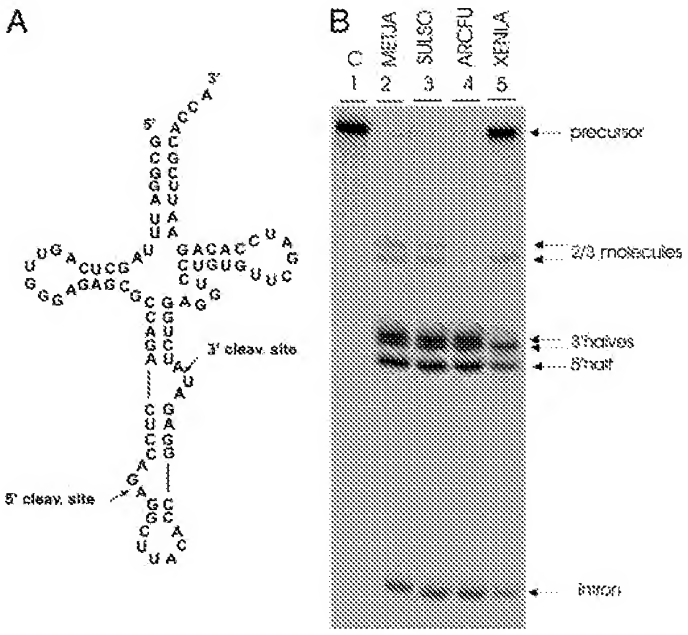

Fig. 4. *In vitro* cleavage of a BHB-containing substrate. (*A*) The pre-tRNA Archeuka was constructed by using two regions derived from yeast pre-tRNA^{Phe} (nucleotides 1–31 and 38–76) joined by a 25-nt insert that corresponds to the BHB motif of archaeal tRNA Trp. (*B*) The pre-tRNA Archeuka was incubated with four different enzymes. The conditions of the reactions are reported in *Materials and Methods*. The cleavage products were analyzed by electrophoresis on 10% polyacrylamide gel containing 29:1 monomer to bis and 8 M Urea, followed by autoradiography. The identification of the reaction products is indicated. Lane 1 is control (C, no enzyme added); lanes 2–5 show the products after incubation with the endonucleases from METJA, SULSO, ARCFU, and XENLA, respectively. The 2/3 molecules are produced by single cleavage. The slowly migrating class corresponds to intron–3′ exon molecules and the fast migrating class to 5′ exon–intron molecules. The heterogeneity of the 3′ halves results from the run-off transcription by T7 RNA polymerase, used to prepare the substrate.

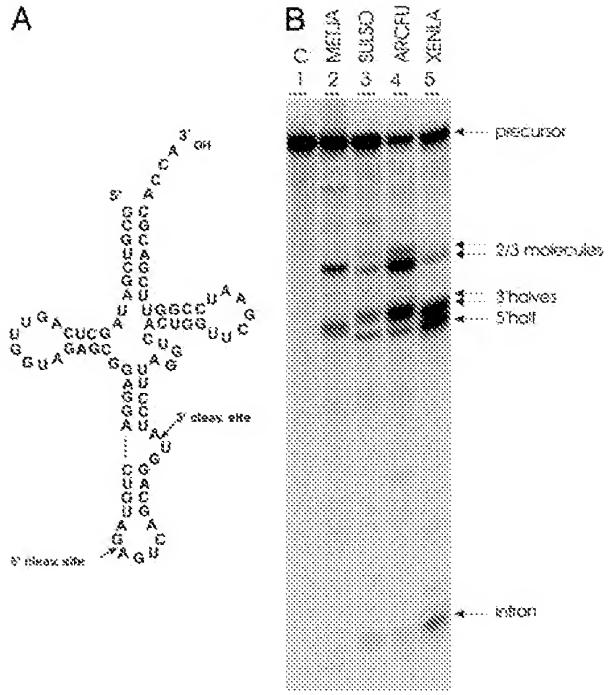

Fig. 5. *In vitro* cleavage of a BHL-containing substrate. (*A*) The pre-tRNA^{Tyr} from *Caenorhabditis elegans*. The synthetic substrate presents a residue change at the 5′ terminus (C to G) required for T7 transcription and a corresponding change (G to C) in the complementary strand. (*B*) The pre-tRNA^{Tyr} was incubated with four different enzymes. The conditions of the reactions and product analysis were the same as described in the legend to Fig. 3.

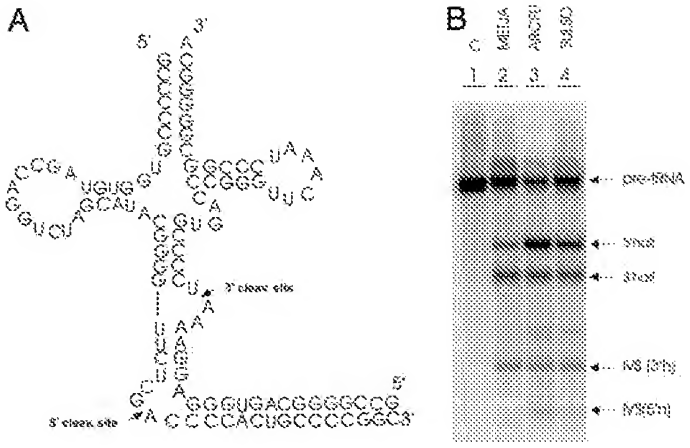

Fig. 6. *In vitro* cleavage of a substrate presenting a trans BHL. (*A*) Proposed structure of the *N. equitans* pre-tRNA^{Glu} formed by annealing the products of two half-genes. The left side of the structure contains the 5′ half-gene, including the 5′ cleavage site in the loop formed by CGAC. The right side of the structure contains the 3′ cleavage site comprising the AAU bulge. The long GC-rich stem at the bottom of the structure is not covalently closed but that does not matter, because it is removed by the splicing endonuclease. (*B*) Cleavage of the pre-tRNA^{Glu} by purified archaeal endonucleases. The substrate was incubated with three different recombinant proteins. The cleavage products were analyzed as described in the legend to Fig. 3. The identification of the reaction products is indicated. Lane 1 is control (C, no enzyme added); lanes 2–4 show the products after incubation with the endonucleases from METJA, ARCFU, and SULSO, respectively. Because the substrate molecule is open at the bottom, the enzymes produce, in addition to 5′ and 3′ halves, two intron fragments. IVS (3′h) is the fragment resulting from cleavage at the 3′ cleavage site, and IVS (5′h) is the fragment resulting from cleavage at the 5′ cleavage site.

now to describe a substrate characterized by a BHL-like intron–exon junction produced by the assembly of two different RNA molecules. *Nanoarchaeum equitans* is a small hyperthermophilic archaeal parasite (20). The *N. equitans* genome contains nine genes that encode tRNA halves; together they account for the missing genes encoding the glutamate, histidine, tryptophan, and initiator methionine transfer RNA species (21). The tRNA sequences are split after the anticodon-adjacent position 37, the normal location of tRNA introns (18). Terminal segments of these half-tRNAs constitute an intervening sequence that includes a 12- to 14-nt, GC-rich RNA duplex formed between the 3′ end of the 5′ tRNA half and the 5′ beginning of the 3′ tRNA half (Fig. 6*A*). The missing *N. equitans* tRNAs reveal the necessity for assembly of two tRNA half-molecules. Randau *et al.* (22) proposed a model based on their discovery of extended reverse complementarity in the intervening sequences: an extended GC-rich duplex in the split intron would facilitate base-pairing of the two halves. This stable duplex in the intron could facilitate folding of the whole tRNA body and promote the cloverleaf structure of the tRNA.

Fig. 6*A* shows a schematic representation of a 5′ tRNA half-gene product (tRNA^{Glu}) and the corresponding 3′ tRNA half-gene product formed in *N. equitans*. The region between the folded tRNA and the intervening RNA duplex at the bottom resembles the BHB. Fig. 6*A* shows that the substrate derived by annealing the RNAs coded by the two tRNA^{Glu} split genes (pre-tRNA^{Glu}) has a 3′ site bulge, a 4-bp helix (with one UG pair), and a 5′ site in a four-base loop. Therefore, we are dealing with a BHL-like motif. Because this structure is located at the position where most archaeal tRNA introns occur, the fusion of the two halves might involve cleavage by the tRNA endonuclease followed by ligation. Here, we demonstrate that tRNA splicing endonucleases can cleave at the expected positions in pre-tRNA^{Glu} (Fig. 6*A*).

We used a representative for each of the three forms of archaeal tRNA endonuclease: the homotetramer from METJA, the homodimer from ARCFU, and the heterotetramer from SULSO. Fig. 6*B*, lanes 3 and 4 shows that pre-tRNA^{Glu} is correctly cleaved by the ARCFU and SULSO enzymes. The products are 5′ and 3′ halves plus the two intron fragments. The identity of the bands was verified by sequencing. The METJA enzyme, on the contrary, requires a strict BHB structure, and it has problems with the recognition and cleavage of the 5′ site (Fig. 6*B*, lane 2). We produced a minisubstrate, lacking the mature domain (5). Consistent with the fact that the archaeal enzymes operate independently of the mature domain, the minisubstrate behaves exactly like the complete pre-tRNA^{Glu} molecule (data not shown).

Discussion

The BHB, renamed hBHBh′ by Marck and Grosjean (10), is a universal substrate (5). It is cleaved twice by all of the characterized tRNA splicing endonucleases, both archaeal and eukaryal. This observation constitutes the main argument to postulate the existence of conserved features among the sites cleaved by the archaeal and the eukaryal enzymes. Certainly conserved is the distance between the active sites (16).

The only available BHB structure, determined by NMR spectroscopy, is that of a 38-nt RNA, derived from *Haloferax volcanii* pre-tRNA Trp (7). The conformation of the two 3-nt bulges is stabilized by stacking interactions between bulge nucleotides and bases in helices H, h, and h′. Both bulges appear on the same minor groove face of the 4-bp helix H. Not all of the archaeal intron–exon junctions can fold into a canonical BHB (hBHBh′) structure. A relaxed hBH or HBh′ motif, including the

constant central 4-bp helix H flanked by one helix (h or h′) with at least 2 bp on each side, is often found (10). We can, therefore, conclude that only helix H and one of the two helices h or h′ are strictly necessary for cleavage in certain Archaea. Because, presumably, the archaeal endonucleases do not contact the mature domain, hBH and HBh′ do not appear different to the enzymes. For this reason we refer collectively to both of them as BHL-like motifs.

Three different forms of tRNA endonuclease can be recognized in Archaea: a homotetramer in some Euryarchaea, a homodimer in other Euryarchaea, and a heteroteramer in the Crenoarchaea and Nanoarchaea. On the basis of the combination of data derived from the study of the phylogenetic distribution of the motifs at the exon–intron junctions (10) and the endonuclease architectures (11) we were led to hypothesize that all three forms of the enzyme can cleave the canonical BHB and that the relaxed BHL-like motif can be cleaved only by the homodimeric (α_2) and the heterotetrameric ($\alpha_2\beta_2$) forms. Our biochemical experiments were designed to explore this hypothetical evolutionary relationship.

Only homodimers or heterotetramers can cleave the BHL-like structures (Figs. 4 and 5). The intron–exon junction motifs and the structures of the enzymes are, therefore, evolutionarily related. Although major questions regarding the origin of tRNA introns are still unanswered, we can speculate that if BHL-like motifs appeared as a consequence of events that modified the BHB motif, it would be necessary to have on hand forms of the enzyme capable of removing the intron correctly. Only those archaeal species that, after gene duplication, present an endonuclease that is either a homodimer or a heterotetramer could process the new substrates. This idea is supported by the fact that some Euryarcheota present a homodimeric endonuclease, but pre-tRNA genes with a BHL-like motif are not encoded in their genomes. It appears that the enzyme specificity for the BHB and BHL-like substrates is the result of adaptation of similar active sites, because the enzymes capable of processing the BHL–like structure are also capable of processing the BHB. This substrate ambiguity is a conspicuous feature that will be evolutionarily exploited in eukaryotic organisms (12, 19, 23).

The intron excision reaction in Eukaryotes is characterized by exquisite dependence on the mature domain (24, 25). The hBH type motif resembles the motif found in most yeast pre-tRNAs presenting introns at 37/38 (18). In this case the bulge is often >3 nt and a conserved base pair between a pyrimidine of the 5′ exon (position 32 in tRNA) and a base in the single-stranded loop of the intron (position 3) is required for correct cleavage of the 3′ splice site (26). The conserved base pair has been called the A-I pair, where A stands for anticodon and I for intron (27). We propose that the relaxed motifs and the consequent ambiguity are a prelude, in the archaeal world, to the loss of autonomy of the BHB-type motifs and the advent of the domination by the mature domain. A common fold, stabilized by the two-conserved-residue signature, characterizes all archaeal endonucleases, despite the existence of the different enzyme forms. The archaeal common fold is not found in the eukaryal enzyme. Again, changes in substrate structure correspond to changes in enzyme structure, according to the paradigms of coevolution (12, 23).

The BHB or the relaxed BHL structures can be formed both in cis and trans. We show that annealing the two tRNA split genes from *N. equitants* produces a substrate for certain archaeal tRNA endonucleases. Because the archaeal tRNA endonuclease does not contact the mature domain of the pre-tRNA, but simply and directly binds to and cleaves the BHB or BHB-like structures, we can expect that transsplicing mediated by the tRNA endonuclease is not restricted to tRNA. In fact, we recently reported that an archaeal endonuclease (from METJA) can catalyze nonspliceosomal mRNA splicing in mouse cells (28).

Note. Kate Calvin, Michelle D. Hall, Fangmin Xu, Song Xue, and Hong Li (29), in agreement with our results, found that the splicing endonuclease from SULSO contains two different subunits and accepts a broad range of substrates.

We thank A. Ferrara for secretarial assistance and G. Di Franco for technical assistance. This work was supported by the Programma Biomolecole per la Salute Umana Ministero dell'Universitá e della Ricerca Scientifica e Tecnologica–Consiglio Nazionale delle Ricerche L95/955%; Fondo Investimenti Ricerca di Base Ministero Instruzione Universitá Ricerca; Progetto Genomica Funzionale L449/97, Ministero Instruzione Universitá Ricerca–Consiglio Nazionale delle Ricerche; Progetto Strategico Tecnologie di Base della Postgenomica, Consiglio Nazionale delle Ricerche; Progetto Strategico Biotecnologie, Ministero dell'Universitá e della Ricerca Scientifica e Tecnologica–Consiglio Nazionale delle Ricerche; Progetto Strategico Genetica Molecolare L449/97 Ministero dell'Universitá e della Ricerca Scientifica; and European Networks of Excellence (EUMORPHIA and MUGEN).

1. Kjems, J., Jensen, J., Olesen, T. & Garrett, R. A. (1989) *Can J. Microbiol.* **35**, 210–214.
2. Kjems, J. & Garrett, R. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 439–443.
3. Kjems, J. & Garrett, R. A. (1988) *Cell* **54**, 693–703.
4. Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J. P., Brosius, J. & Huttenhofer, A. (2002) *Nucleic Acids Res.* **30**, 921–930.
5. Fabbri, S., Fruscoloni, P., Bufardeci, E., Di Nicola Negri, E., Baldi, M. I., Attardi, D. G., Mattoccia, E. & Tocchini-Valentini, G. P. (1998) *Science* **280**, 284–286.
6. Thompson, L. D. & Daniels, C. J. (1990) *J. Biol. Chem.* **265**, 18104–18111.
7. Diener, J. L. & Moore, P. B. (1998) *Mol. Cell* **1**, 883–894.
8. Kleman-Leyer, K., Armbruster, D. W. & Daniels, C. J. (1997) *Cell* **89**, 839–847.
9. Lykke-Andersen, J. & Garrett, R. A. (1997) *EMBO J.* **16**, 6290–6300.
10. Marck, C. & Grosjean, H. (2003) *RNA* **9**, 1516–1531.
11. Tocchini-Valentini, G. D., Fruscoloni, P. & Tocchini-Valentini, G. P. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 8933–8938.
12. Jensen, R. A. (1976) *Annu. Rev. Microbiol.* **30**, 409–425.
13. Hughes, A. L. & Friedman, R. (2005) *Evol. Dev.* **7**, 196–200.
14. Fribourg, S., Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A. & Moras, D. (2001) *J. Mol. Biol.* **306**, 363–373.
15. Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E. & Tocchini-Valentini, G. P. (1997) *Cell* **89**, 859–866.
16. Li, H., Trotta, C. R. & Abelson, J. (1998) *Science* **280**, 279–284.
17. Li, H. & Abelson, J. (2000) *J. Mol. Biol.* **302**, 639–648.
18. Abelson, J., Trotta, C. R. & Li, H. (1998) *J. Biol. Chem.* **273**, 12685–12688.
19. Fruscoloni, P., Baldi, M. I. & Tocchini-Valentini, G. P. (2001) *EMBO Rep.* **2**, 217–221.
20. Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C. & Stetter, K. O. (2002) *Nature* **417**, 63–67.
21. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 12984–12988.
22. Randau, L., Munch, R., Hohn, M. J., Jahn, D. & Soll, D. (2005) *Nature* **433**, 537–541.
23. Hughes, A. L. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 8791–8792.
24. Reyes, V. M. & Abelson, J. (1988) *Cell* **55**, 719–730.
25. Mattoccia, E., Baldi, I. M., Gandini-Attardi, D., Ciafre, S. & Tocchini-Valentini, G. P. (1988) *Cell* **55**, 731–738.
26. Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S. & Tocchini-Valentini, G. P. (1992) *Science* **255**, 1404–1408.
27. Abelson, J. (1992) *Science* **255**, 1390.
28. Deidda, G., Rossi, N. & Tocchini-Valentini, G. P. (2003) *Nat. Biotechnol.* **21**, 1499–1504.
29. Calvin, K., Hall, M. D., Xu, F., Xue, S. & Li, H. (2005) *J. Mol. Biol.*, in press.

The dawn of dominance by the mature domain in tRNA splicing

Giuseppe D. Tocchini-Valentini*†, Paolo Fruscoloni*, and Glauco P. Tocchini-Valentini*‡

*Istituto di Biologia Cellulare, Consiglio Nazionale delle Ricerche, Campus A, Buzzati-Traverso, Via Ramarini 32, Monterotondo Scalo, 00016 Rome, Italy; and †Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche, Via Abbiategrasso 207, I-27100 Pavia, Italy

Communicated by Robert Haselkorn, University of Chicago, Chicago, IL, June 13, 2007 (received for review May 23, 2007)

The relationship between enzyme architecture and substrate specificity among archaeal pre-tRNA splicing endonucleases has been investigated more deeply, by using biochemical assays and model building. The enzyme from *Archeoglobus fulgidus* (AF) is particularly interesting: it cleaves the bulge–helix–bulge target without requiring the mature tRNA domain, but, when the target is a bulge–helix–loop, the mature domain is required. A model of AF based on its electrostatic potential shows three polar patches interacting with the pre-tRNA substrate. A simple deletion mutant of the AF endonuclease lacking two of the three polar patches no longer cleaves the bulge–helix–loop substrate with or without the mature domain. This single deletion shows a possible path for the evolution of eukaryal splicing endonucleases from the archaeal enzyme.

molecular evolution | RNA–protein interactions | tRNA endonucleases

Accuracy in tRNA splicing is essential for the formation of functional tRNAs and therefore for gene expression. In Bacteria, pre-tRNA introns are self-splicing group I introns, and the splicing mechanism is autocatalytic. In Eukarya, tRNA introns are small and invariably interrupt the anticodon loop one base 3′ to the anticodon. In Archaea, the introns are also small and often reside in the same location as eukaryal tRNA introns, but not always (1, 2). In both Eukarya and Archaea, the specificity for recognition of the pre-tRNA resides in the endonucleases (3–5). These enzymes remove the intron by making two independent endonucleolytic cleavages. It is generally accepted that the archaeal enzymes act without any reference to the mature domain but instead recognize specific structures that define the intron–exon boundaries (6–8). By contrast, the eukaryal enzyme normally acts in a mature-domain-dependent mode (3, 4).

The exact way in which the eukaryal endonucleases recognize the precursors has yet to be determined, but many RNA–protein interactions are presumably required. Previous studies have shown the role of the three-dimensional structure and of specific invariant bases of the mature domain in recognition and binding by the enzyme (3, 4). The importance of structural regularities in the precursors suggested an anchor-and-measure mechanism in which the endonuclease binds to one or more reference sites in the mature domain, which are common to all pre-tRNAs, and measures the distance to the equivalently positioned intron–exon junctions. This hypothesis was supported by experiments that involved the engineering of changes in the distance that separate the mature domain from the splice sites. These alterations changed the size of the intron in a predictable way: the insertion of 1 bp in the anticodon stem increased the size of the intron by two bases, one at each end, and the insertion of 2 bp in the anticodon stem increased the size of the intron by 4 nt (3, 4). These striking results, which had been observed in specific instances, were generalized and taken to mean that there are no important recognition elements at the intron–exon boundaries and that the enzyme binds only to the mature domain. Using a large number of constructs and *in vitro* selection techniques, we later found evidence that pre-tRNA introns also contribute to the specificity of splice site recognition (9, 10). There are also positions in the mature domain that participate in the splicing reaction, although they are not occupied by the same bases in all pre-tRNAs. We called these positions "cardinal positions" (CPs) (9). The eukaryal endonuclease is able to cut a precursor at the 3′ site if a base in the pre-tRNA single-stranded loop of the intron is allowed to form an anticodon-intron (A-I) base pair with the base of the 5′ exon situated at the position immediately following the anticodon stem (position 32 in yeast pre-tRNA^{Phe}). A recently observed "A-minor" interaction between the first 3′ exon nucleotide and the A-I base pair provides the structural basis for their interdependence (11). These observations suggested that important recognition elements also lie at the intron–exon boundaries.

The mature domain dominates the scenario of eukaryal tRNA splicing. There is, however, a single striking exception: a specific structure consisting of two 3-nt bulges separated by a 4-bp helix, the so called bulge–helix–bulge (BHB) motif, has been shown to be recognized and cleaved *in vitro* by the eukaryal enzyme without any reference to the mature domain (12, 13). This finding presents interesting evolutionary implications. The BHB pre-tRNA is a universal substrate and is also cleaved by the archaeal enzymes. As a matter of fact, many intron-containing archaeal pre-tRNAs are characterized by a BHB constructed from the intron and from exonic sequences. The archaeal endonucleases, in general, remove the intron in a mature-domain-independent mode (14). It is because of this independence that the intron in archaeal pre-tRNAs can be located in different positions relative to the mature domain. Several archaeal pre-tRNAs contain relaxed forms of the BHB motif, one being the bulge–helix–loop (BHL), that consists of a single 3-nt bulge and an internal loop, separated by a 4-bp-helix (15). Some archaeal enzymes, such as the enzyme from *Sulfolobus solfataricus* (SS), can cleave the BHL substrate in a mature-domain-independent way. We reasoned, however, that, because the relaxed forms like the BHL resemble the motif found in most yeast pre-tRNAs (presenting, as a rule, introns located one base 3′ to the anticodon), they could represent a preliminary stage in the advent of the dominance of the tRNA splicing by the mature domain (15).

At a certain stage, an enzyme could have evolved the capability of cleaving the BHL substrate only in mature-domain-dependent mode, while still capable of cleaving the BHB substrate in a mature-domain-independent mode. Such an enzyme would be characterized by surfaces providing specific interaction with the mature domain. In this article, we describe results of *in vitro* cleavage assays performed using a variety of substrate variants and the endonuclease from the archaeon *Archeoglobus fulgidus* (AF). This enzyme cleaves in a mature-domain-independent mode the BHB substrate but not the BHL substrate. In this respect, the AF enzyme resembles the eukaryal enzymes. An electrostatic potential

Author contributions: G.D.T.-V., P.F., and G.P.T.-V. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Abbreviations: NAF, N-terminal repeat; CAF, C-terminal repeat; BHB, bulge–helix–bulge; BHL, bulge–helix–loop; SS, *Sulfolobus solfataricus*; AF, *Archeoglobus fulgidus*; MJ, *Methanocaldococcus jannaschii*.

‡To whom correspondence should be addressed. E-mail: gtocchini@ibc.cnr.it.

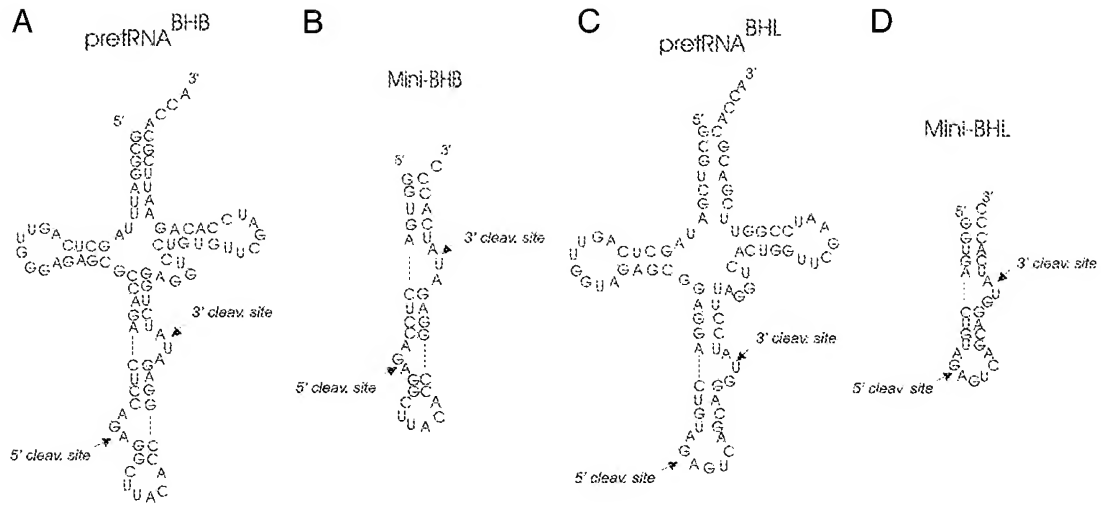

Fig. 1. Substrates for the *in vitro* cleavage assays. (*A*) PretRNA^{BHB} consists of two regions derived from yeast pre-tRNA^{PHE} (nucleotides 1–31 and 38–76), joined by a 25-nt insert that corresponds to the BHB motif of archaeal tRNA^{Trp}. (*B*) Mini-BHB is the BHB motif of pre-tRNA^{BHB}. (*C*) PretRNA^{BHL} corresponds to pre-tRNA^{Tyr} from *Caenorhabditis elegans*. The synthetic substrate presents a residue change at the 5′ terminus (C to G) required for T7 transcription and a corresponding change (G to C) in the complementary strand. (*D*) Mini-BHL is the BHL motif of pre-tRNA^{BHL}. The base substitutions in the mini substrates were introduced to optimize T7 RNA polymerase transcription (5).

calculation (16) applied to the AF enzyme structure permits the visualization, on the face of the protein that interacts with pre-tRNA, of positively charged patches, presumably involved in the interaction with specific parts of the substrate, including the mature domain.

Results

The AF Enzyme Cleaves the BHL Motif Substrate in a Mature-Domain-Dependent Mode. We reasoned that, if in the archaeal world tRNA endonucleases exist that interact with the mature domain following a strategy similar to that adopted by the eukaryal enzymes, then in the corresponding organism there should exist pre-tRNAs presenting the intron located only at the canonical position in the anticodon loop region between nucleotides 37 and 38. Interestingly, two independent bioinformatic analyses have reached the same conclusion: the Euryarchaea encoding a homodimeric endonuclease present pre-tRNAs with the intron located exclusively at this canonical position (14). Do homodimeric enzymes, like the AF endonuclease, occupy a critical position in the evolutionary route from the archaeal to the eukaryal system?

To investigate this possibility, we used two different pre-tRNA substrates for the *in vitro* cleavage assay. The first substrate pre-tRNA^{BHB} presents a BHB motif whereas the second pre-tRNA^{BHL} presents a BHL (Fig. 1*A* and *C*). Each substrate was incubated with three different pre-tRNA endonucleases from Archaea, representing each of the three different architectures: the homotetramer (α_4) of *Methanocaldococcus jannaschii* (MJ), the homodimer (α_2) of AF, and the heterotetramer ($\alpha_2\,\beta_2$) of SS. All of the enzymes cleaved the BHB substrate (Fig. 2*C*). On the contrary, there are differences in the processing of the BHL substrate. The homotetrameric enzyme from MJ did not cleave the BHL. The heterotetrameric enzyme from SS and the homodimeric enzyme from AF cleaved the BHL substrate, which is, therefore, stable at the assay conditions (65°C) (Fig. 2*D*). These findings confirm results previously obtained by us (15). It seems, therefore, that the enzyme specificity for the BHB and BHL substrates is the result of adaptation of similar active sites, because the enzymes capable of processing the BHL pre-tRNA are also capable of processing the BHB. The BHB is a universal substrate. The question we address next is whether the ability to cleave the relaxed motifs is a prelude, in the archaeal world, to the advent of the dominance by the mature domain in the Eukarya.

We asked first whether each of the three archaeal enzymes, representing the different architectures (17), can operate independently of the mature domain on minisubstrates that correspond respectively to the BHB (mini-BHB) (Fig. 1*B*) or to the BHL (mini-BHL) (Fig. 1*D*) and lack the mature domain. Fig. 2*A* shows that the intron was correctly removed from the mini–BHB by all three enzymes whereas the mini-BHL was effectively cleaved only by the heterotetrameric SS endonuclease (Fig. 2*B*). The active sites of the enzyme from SS therefore allow the BHL to perform autonomously. We expected that the homotetrameric enzyme from MJ would not cleave this substrate. But to our surprise, the homodimeric AF enzyme also failed to cleave the mini-BHL (Fig. 2*B*). Inefficient cleavage of the 3′ site could occasionally be observed (see Fig. 5*B*). This striking observation is in contrast to our observation for the full-length pre-tRNA^{BHL} (Fig. 2*D*). These results suggest that the homodimeric enzyme of AF, like the eukaryal enzymes, requires the mature domain for cleavage. These observations led us to attempt to visualize a footprint of the full-length pre-tRNA, on the surface of the AF enzyme, a footprint of the full-length pre-tRNA.

The AF Enzyme Presents RNA Binding Sites Involved in Specific Interactions with Different tRNA Domains. We first calculated an electrostatic potential using the MJ endonuclease structure determined by x-ray diffraction (18). This enzyme presents a face interacting with tRNA that is uniformly positively charged (Fig. 3*A*). Instead, the AF enzyme surface charge, calculated in the same way, omitting the RNA substrate, clearly shows three distinct positively charged polar patches (11) (Fig. 3*B*). The central patch is formed by residues belonging to the two catalytic repeats; it includes the two catalytic sites and is directly involved in the interaction with the BHB. The two flanking patches are identical and symmetrically placed with respect to one another. These two peripheral patches are formed by 11 basic residues located in the N-terminal domain of the AF protein. Four belong to helix 1 (Lys-12, Lys-14, Arg-18, and Arg-19); two belong to to helix 4 (Arg-74 and Arg-76); and two are on loop 7 (Lys-90 and Lys-91) (11).

To confirm that we are dealing with three discrete patches and not a continuous elongated one, we considered the fact that each subunit of the homodimeric AF enzyme contains two similar repeating domains that are homologous to the subunit structure of the homotetrameric enzyme from MJ: the C-terminal repeat

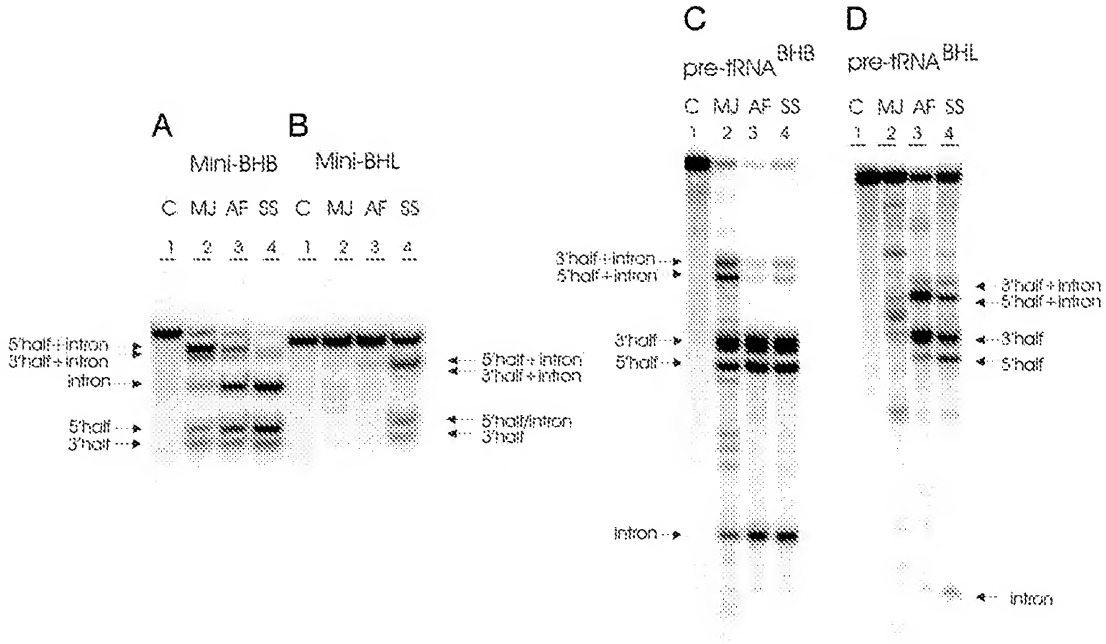

Fig. 2. *In vitro* cleavage of BHB- and BHL-containing substrates. (*A*) The mini–BHB substrate was incubated with three different enzymes. (*B*) The mini–BHL substrate was incubated with three different enzymes. (*C*) Pre-tRNA^{BHB} was incubated with three different enzymes. (*D*) PretRNA^{BHL} was incubated with three different enzymes. The conditions of the reactions have been reported (15). The cleavage products were analyzed by electrophoresis on 10% polyacrylamide gel containing 29:1 monomer to bis and 8 M urea, followed by autoradiography. The identification of the reaction products is indicated. Lane 1 contains the control (C, no enzyme added). Lanes 2–4 show the products after incubation with the endonucleases from MJ, AF, and SS, respectively. The 2/3 molecules are produced by single cleavage.

(CAF) and the N-terminal repeat (NAF). We designed a mutant enzyme that comprises a deletion of the NAF domain with the exception of the last two beta strands that comprise loop 10 at the C terminus, known to be important for the assembly of the enzyme (Fig. 4*A*). According to our expectations, the mutant enzyme (ΔAF) lacks the two peripheral patches, but it should still present all of the residues directly interacting with the BHB (Fig. 4*B*). To determine whether this truncated enzyme can assemble into an active form, we overexpressed and purified it as described in *Materials and Methods*. Fig. 5 *A* and *C* shows that the truncated enzyme is capable of cleaving the BHB in the presence or absence of a tRNA mature domain, but it is no longer capable of cleaving the BHL substrate, even in the presence of the mature domain (Fig. 5 *B* and *C*). We conclude that the positively charged peripheral patches interact with a set of recognition elements that are presented by the mature domain of AF tRNA precursors.

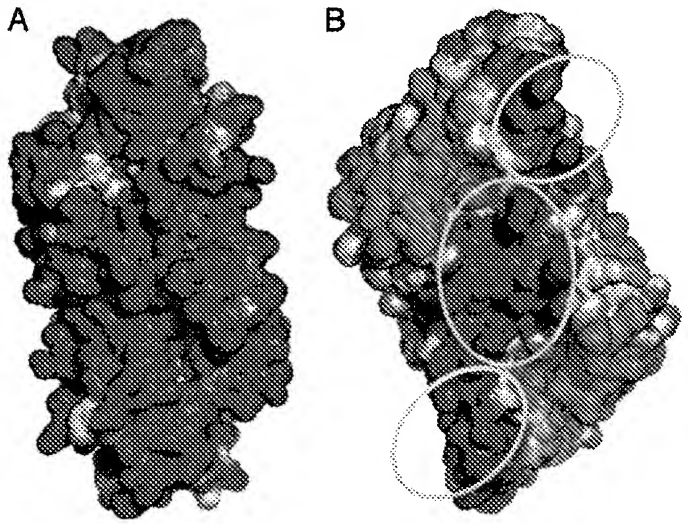

Fig. 3. Electrostatic surface potentials of MJ (*A*) and AF (*B*) endonucleases. Positively charged regions are shown in blue, the neutral regions are shown in white, and the negatively charged regions are shown in red. The green circle encompasses the catalytic site, whereas the cyan circles enclose the two putative sites recognizing the tRNA mature domain. The figure was generated with PyMOL (http://pymol.sourceforge.net), and the electrostatic properties were generated by using the APBS (Adaptive Poisson-Boltzman Solver) software package (16).

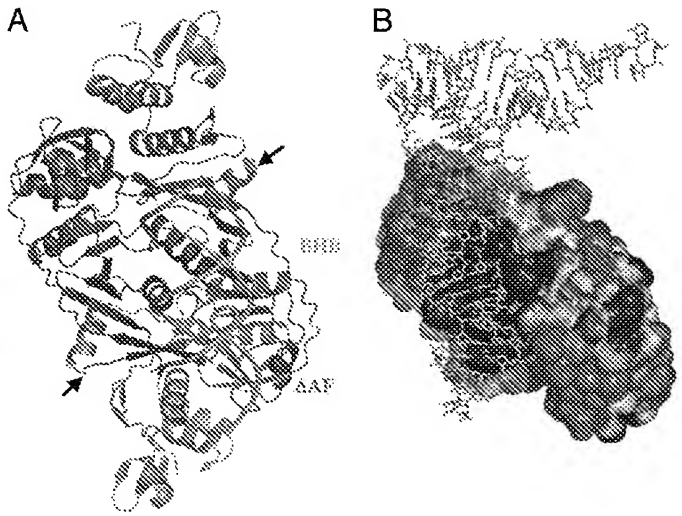

Fig. 4. Structure of ΔAF enzyme and model of the interaction with its substrate. (*A*) Cartoon representation of the parts deleted from AF endonuclease (gray). The arrows indicate the new N termini of the subunits. The BHB substrate is shown in green. (*B*) Electrostatic surface potential of ΔAF. Positively charged regions are shown in blue, the neutral regions are shown in white, and the negatively charged regions are shown in red. The pre-tRNA model is shown in cyan.

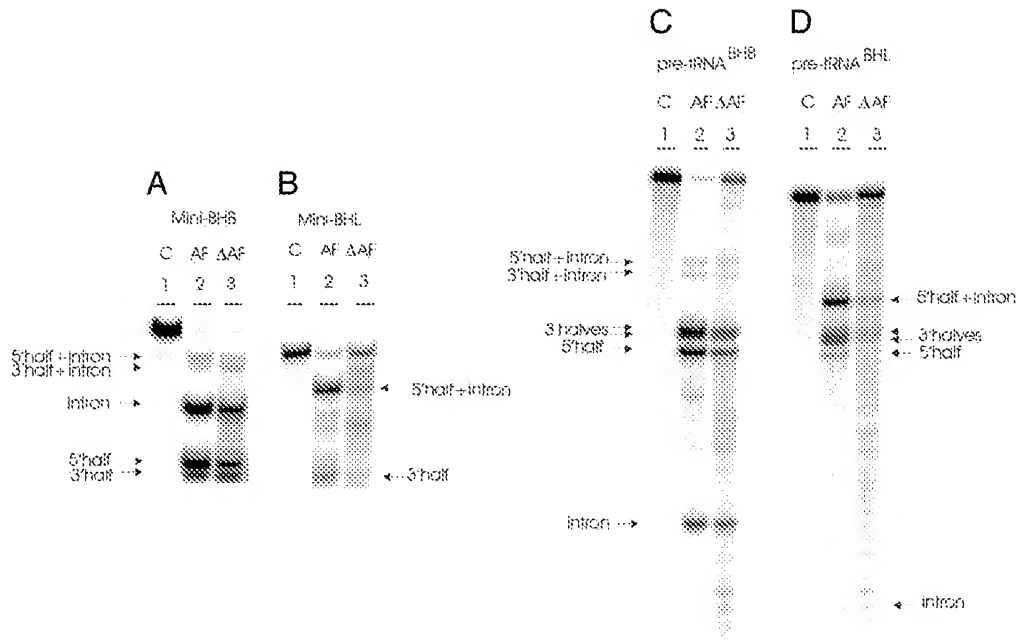

Fig. 5. *In vitro* cleavage of BHB and BHL containing substrates by the ΔAF enzyme. (*A*) The mini-BHB substrate was incubated with the AF and ΔAF enzymes. (*B*) The mini-BHL substrate was incubated with the AF and ΔAF enzymes. (*C*) The pre-tRNA^{BHB} substrate was incubated with the AF and ΔAF enzymes. (*D*) The pre-tRNA^{BHL} substrate was incubated with the AF and ΔAF enzymes. The conditions of the reactions have been reported (12). The cleavage products were analyzed by electrophoresis on 10% polyacrylamide gel containing 29:1 monomer to bis and 8 M urea, followed by autoradiography. The identification of the reaction products is indicated. Lane 1 contains the control (C, no enzyme added). Lanes 2 and 3 show the products after incubation with the endonucleases from AF and ΔAF, respectively. The 2/3 molecules are produced by single cleavage.

Discussion

In Archaea, there are three distinct architectures of tRNA endonuclease, the enzyme responsible for the excision of the intron from pre-tRNA. The crystal structures of the enzymes from two Euryarcheota, AF and MJ, are known (18, 19). The AF enzyme is a homodimer (α_2). Each subunit comprises two similar repeats, but the same repeat performs a different function in each of the two subunits. The NAF acts to stabilize the dimer, whereas the CAF is the catalytic domain. The MJ enzyme, a homotetramer (α_4), is characterized by a similar functional subdivision. Two of the subunits play a catalytic role, whereas the other two have a structural role. We found that the Crenarcheote SS presents two genes homologous to the one coding for the MJ protein. One of the two genes is specialized for encoding the catalytic subunit and the other for encoding the subunit that maintains the structure of the heterotetramer ($\alpha_2\beta_2$). Presumably, at the origin, in Archaea the situation was similar to that in MJ, with a single gene coding for a single protein able to serve both as catalytic or structural subunit. During evolution in AF and in SS, the single gene duplicated. In AF the two duplicated genes fused and subfunctionalization occurred between the two halves of the fused protein: the C-terminal half (CAF) retained the catalytic function and the N-terminal half (NAF) specialized as the structural subunit. In SS, one of the duplicated genes has specialized for encoding the catalytic and the other for encoding the structural subunit. Both subunits are necessary for substrate cleavage by the heterotetramer (15, 17, 20, 21).

The protagonist of this paper is the AF enzyme. The homodimer cleaves pre-tRNA^{BHB}, pre-tRNA^{BHL}, and mini-BHB, but not mini-BHL. The AF enzyme, therefore, can cleave the BHL only in a mature-domain-dependent mode. Because mature-domain dependence for cleavage of non-BHB substrates is a main feature of the eukaryotic enzymes, we observe in AF the dawn of dominance of the mature domain in tRNA splicing. In AF and in all other Euryarchaea encoding homodimeric endonucleases, as in Eukarya, the intron is located at the canonical position in the anticodon loop region between nucleotides 37 and 38. The latter feature is a necessary requirement for mature-domain dependence, because it assures a fixed geometry relative to the mature domain for all pre-tRNAs. Three RNA-binding sites, one central and two peripheral, characterize the surface of the AF enzyme that interacts with the substrate (Fig. 3*B*). The central site contains the two symmetric catalytic triads and presumably interacts with the BHB and its relaxed forms, like the BHL. The two flanking patches are identical and symmetrically placed and are presumably involved in the interaction with the mature domain. They function one at a time, because the enzyme can accommodate only one tRNA molecule. In an experiment that attempted to reverse the proposed evolutionary process, we deleted a portion of the N-terminal domain of the AF enzyme; the resulting ΔAF mutant should lack the two peripheral RNA binding regions (Fig. 4 *A* and *B*). As predicted, ΔAF is unable to perform in a mature-domain-dependent mode. It can cleave only pre-tRNA^{BHB} and mini-BHB; it does not cleave pre-tRNA^{BHL} or, of course, the mini-BHL (Fig. 5).

Interaction of a Pre-tRNA^{BHB} with the AF Enzyme. We built a model using the recently solved structure of the AF enzyme in complex with a minimal BHB substrate, generously provided by H. Li (11). Interestingly, the asymmetric unit of the co-crystal contains two molecules of RNA–enzyme complex and the analysis of the crystal packing made it possible to see that each enzyme, besides interacting with its substrate through the BHB structure, also interacts with the A-helix belonging to the substrate of a symmetry-related enzyme molecule (Fig. 6*A*). These interactions involve the side chains of Lys-90 and Lys-91, which belong to the group of residues forming the above described peripheral positive patch (Fig. 3*B*). Almost certainly, the observed RNA–protein contacts are the result of different packing constraints in the crystal lattice and only mimic the interaction with a cognate mature-domain substrate in solution, but they clearly suggest that a site can be available to host the tRNA mature body as we propose in our model. Because of the symmetry

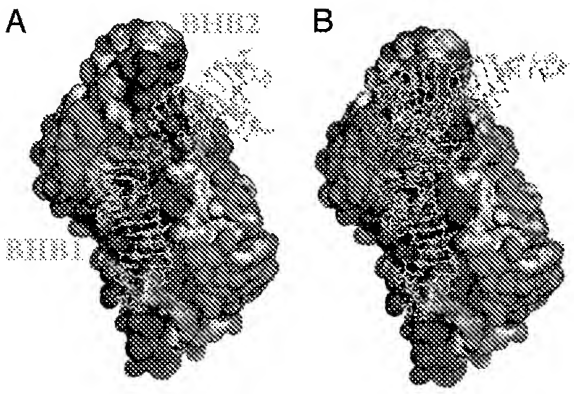

Fig. 6. Model of the interactions between the AF enzyme and the substrates. Electrostatic surface potential of AF bound either to two BHB RNA molecules (shown in green) obtained from the crystallographic structure (11) (*A*) or to a model of pre-tRNA (shown in cyan) (*B*).

of the catalytic site, the enzyme can accommodate one tRNA molecule at a time, interacting through one or another patch. We superimposed, using the program O, the 6-bp stem of the anticodon arm of the $tRNA^{Phe}$ to the corresponding 6-bp stem of the BHB minimal substrate. The superimposed pre-tRNA mature domain faces one of the peripheral patches of the enzyme with the variable pocket and T-arm-acceptor stem helix (Fig. 6*B*). The enzyme can approach the tRNA only by using this strategy; introns are also present in class II tRNAs ($tRNA^{Leu}_{CAA}$ in AF). These tRNAs present a long extra arm that could otherwise impede, by steric hindrance, interaction with an enzyme approaching from the opposite side. Moreover, the T-arm-acceptor coaxial A-helix presents its deep and narrow major groove to the enzyme so that the latter can interact only with the phosphate backbone, thereby permitting recognition of the tRNA substrate that is not sequence specific.

From Archaea to Eukarya. The salient features that characterize the AF enzyme are the result of subfunctionalization and creation of the binding sites for elements of the mature domain of the pre-tRNA. As far as subfunctionalization is concerned, presumably, the extreme environments that characterize the habitat of Archaea is not ideal for α_4 architectures, where a single polypeptide plays two very different roles. It is not clear whether drift alone or positive selection or both had a role in subfunctionalization.

We should not forget, however, that subfunctionalization did not occur only in homodimeric ($\alpha\beta$) enzymes, like AF, but also in heterotetrameric ($\alpha_2\beta_2$) enzymes like SS. The fact, therefore, that it occurred twice independently, and by different routes in the same gene, argues in favor of positive selection. The acquisition of the three RNA-binding sites points to the extraordinary and spectacular series of events that saw evolution engraving on the surface of the AF enzyme the overlapping portraits of two pre-tRNAs. The ΔAF mutant tells us that we can reverse the evolutionary process by deleting a sequence of amino acids from the N-terminal part of

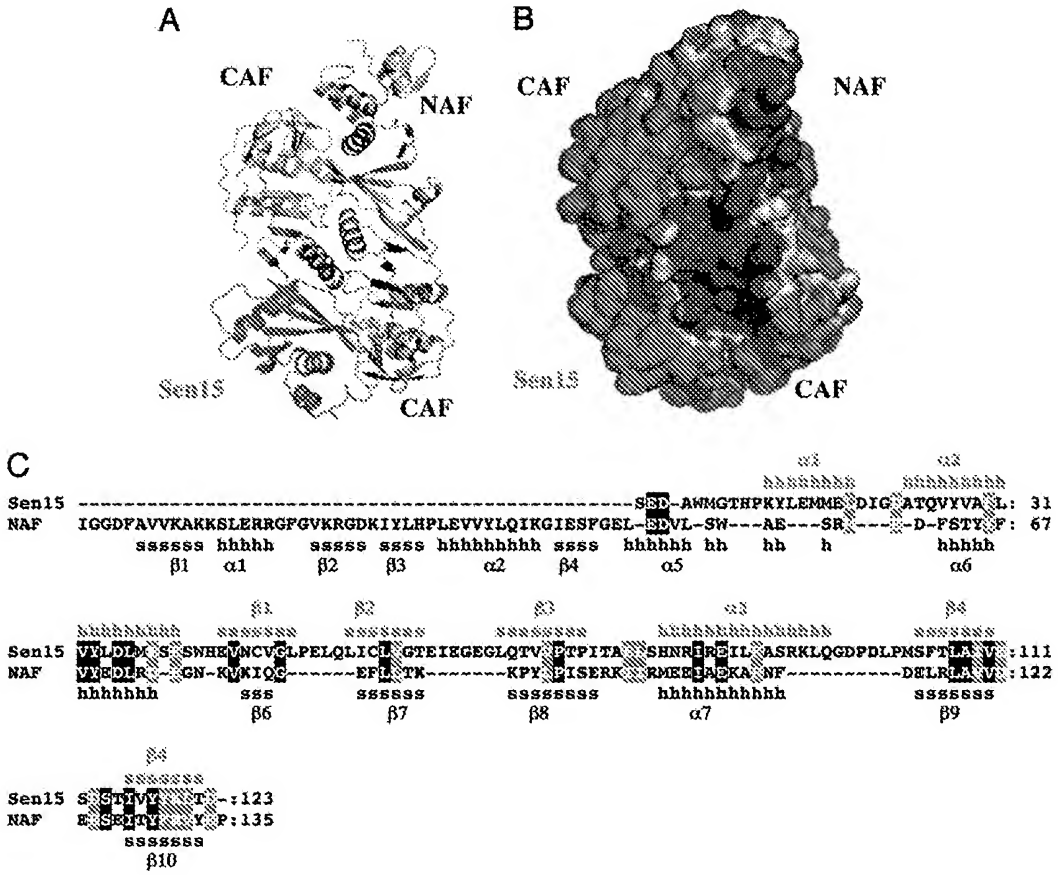

Fig. 7. Model of a chimeric enzyme presenting two CAFs, one NAF, and the human Sen15. (*A*) Representation of the model chimeric enzyme. The archaeal parts are shown in gray, and the human part is shown in red. (*B*) Electrostatic surface potential of the model chimeric enzyme. Positively charged regions are shown in blue, neutral regions are shown in white, and negatively charged regions are shown in red. (*C*) Primary sequences and secondary structures of the NAF and of the human Sen15. Secondary structure elements, as determined by the crystallographic structure of AF (19) and the NMR structure of Sen15 (24), are shown in black and red, respectively. The "s" and β indicate β-strands, and "h" and α indicate α-helices. Black columns indicate residues that are conserved; gray columns indicate residues that are similar.

the NAF. The truncated enzyme loses the ability to act in a mature-domain-dependent mode. If we compare Fig. 4*B* with Fig. 6*B*, we see that ΔAF completely lacks the peripheral sites that interact, in the wild-type enzyme, with the mature domain.

In the intact AF enzyme, the two active sites are completely interchangeable. Depending on which peripheral binding site is used by the pre-tRNA substrate, a given active site will cleave at the 5′ or at the 3′ junction. In the case of the yeast tRNA splicing endonuclease, the situation is radically different. The eukaryal enzyme is a heterotetramer (α, β, γ, δ) (22). The four subunits (Sen15, Sen2, Sen34, and Sen54 in yeast) contribute to the formation of two composite active sites that, because of specific interactions with the mature domain, interact rigidly and specifically either with the 5′ or the 3′ cleavage sites (23). If we wanted, therefore, to modify the AF enzyme to make it more similar to its eukaryal counterpart, we should think of ways to eliminate one of the two peripheral patches. In this way, the active sites would be rigidly assigned to one or the other of the two cleavage sites.

The ΔAF result teaches us that the peripheral sites comprise a sequence of amino acids located in the N-terminal part of the NAF. We observed that the sequence of the eukaryal subunit Sen15 does not contain the catalytic triad (22) and that therefore it must be a structural subunit. Alignment of the Sen15 and NAF sequences shows that the eukaryal peptide lacks a segment corresponding to the N-terminal part of NAF, which constitute the peripheral site, but is homologous to the C-terminal sequence of the AF enzyme (Fig. 7*C*). Because the NMR structure of the human Sen15 (24) is available, we aligned the atomic structures of the Sen15 monomer with NAF using the program VMD (25). The alignment clearly shows that Sen15 and the C-terminal portion of NAF share a common fold (Fig. 7*A*). On the basis of these observations, we constructed a model of a chimeric enzyme comprising two CAFs, one NAF, and the human Sen15 substituting for the other NAF. Electrostatic potential calculations show that this model chimeric enzyme has lost one of the peripheral patches and, like a eukaryal enzyme, is characterized by active sites rigidly assigned to either the 5′ or the 3′ cleavage sites of the substrate (Fig. 7*B*).

We expect that, if the route that we hypothesize was really followed in evolution, the active sites would drift, obviously conserving the catalytic triad and the ability to cleave the BHB substrate, adapting to the specific intron–exon boundaries. Clearly, much remains to be done in terms of defining the parts of the mature domain that are required for correct positioning of the pre-tRNA on the enzyme surface. What we have established is that at least one current archeon, AF, retains the memory of the first step on the path to the eukaryal system.

Materials and Methods

Expression and Purification of the Protein Constructs. The construction of the vectors coding for the genes of the endonucleases from AF, SS, and MJ and the purification of the enzymes are described elsewhere (15, 17). The gene coding for the ΔAF construct was obtain by PCR-amplification by using as a template the previously cloned AF gene and the following DNA primers: (*i*) 5′GGAATTCCATATGGACGAGTTAAGGCTTGCTGTCG and (*ii*) 5′CGCGGATCCTCAAACCTTAACCCTCTCAAAGC. The two primers were designed to obtain an amplified fragment presenting an NdeI site upstream of the gene and a BamHI site downstream. After digest, the fragment was cloned into pET28b (Novagen, Madison, WI). The correct clones were verified by DNA sequencing. The protein was overexpressed as hexa-histidine-tagged forms (pET28b) in *Escherichia coli* Rosetta (Novagen). Cells were grown in 1-liter cultures of Terrific Broth broth at 37°C in the presence of 30 μg/ml kanamycin (pET28) with the addition of 30 μg/ml chloramphenicol. The purification consisted of a metal affinity column as a first step, followed by gel filtration using an analogous strategy already described for the other enzymes (17). The purity of the enzyme was assessed by Coomassie blue staining of SDS polyacrylamide gels.

***In Vitro* RNA Synthesis and Pre-tRNA Splicing Reactions.** DNA templates were produced as described (17). T7 RNA polymerase transcription reactions were carried out following the conditions of the Ambion (Austin, TX) T7-Megashortscript kit. [α-^{32}P]UTP (800 Ci/mmol; Amersham Pharmacia) (1 Ci = 37 GBq) was included to radiolabel the RNA transcripts. All reactions were purified by electrophoresis on a 10% denaturing polyacrylamide gel. RNA products corresponding to the correct size were eluted, phenol was extracted, and ethanol was precipitated.

tRNA splicing reactions were typically performed in 25 mM Tris·HCl (pH 7.5), 5 mM $MgCl_2$, 100 mM NaCl, and 10% glycerol, and included 20 fmol of tRNA precursor substrates. Purified splicing endonucleases were added and incubated at 65°C for 1 h. The reactions were stopped by phenol extraction, and ethanol was precipitated and separated on 10% denaturing polyacrylamide gels. Cleavage products were visualized by analysis on a Molecular Dynamics model Storm 860 PhosphorImager by using ImageQuant software, version 4.

This article is dedicated to the memory of our beloved friend and colleague Domenica "Nica" Gandini-Attardi, prematurely deceased, who was very helpful for continuous encouragement and critical reading of the manuscript. We thank Hong Li for providing the coordinates of the AF splicing endonuclease bound with an RNA BHB substrate (Protein Data Bank ID code 2GJW), A. Ferrara and T. Cuccurullo for secretarial assistance, and G. Di Franco for technical assistance. This work was supported by Italian Ministry of Research FIRB (G. Armenise-Harvard Foundation, Italia-Canada and Idee Progett. 2005) and SVIFASTA grants and European FP6 contracts (MUGEN, EURASNET, and EUMODIC).

1. Abelson J, Trotta CR, Li H (1998) *J Biol Chem* 273:12685–12688.
2. Belfort M, Weiner A (1997) *Cell* 89:1003–1006.
3. Reyes VM, Abelson J (1988) *Cell* 55:719–730.
4. Mattoccia E, Baldi IM, Gandini-Attardi D, Ciafre S, Tocchini-Valentini GP (1988) *Cell* 55:731–738.
5. Di Nicola Negri E, Fabbri S, Bufardeci E, Baldi MI, Gandini Attardi D, Mattoccia E, Tocchini-Valentini GP (1997) *Cell* 89:859–866.
6. Palmer JR, Nieuwlandt DT, Daniels CJ (1994) *J Bacteriol* 176:3820–3823.
7. Kleman-Leyer K, Armbruster DW, Daniels CJ (1997) *Cell* 89:839–847.
8. Lykke-Andersen J, Garrett RA (1997) *EMBO J* 16:6290–6300.
9. Baldi MI, Mattoccia E, Bufardeci E, Fabbri S, Tocchini-Valentini GP (1992) *Science* 255:1404–1408.
10. Bufardeci E, Fabbri S, Baldi MI, Mattoccia E, Tocchini-Valentini GP (1993) *EMBO J* 12:4697–4704.
11. Xue S, Calvin K, Li H (2006) *Science* 312:906–910.
12. Fruscoloni P, Baldi MI, Tocchini-Valentini GP (2001) *EMBO Rep* 2:217–221.
13. Fabbri S, Fruscoloni P, Bufardeci E, Di Nicola Negri E, Baldi MI, Attardi DG, Mattoccia E, Tocchini-Valentini GP (1998) *Science* 280:284–286.
14. Marck C, Grosjean H (2003) *RNA* 9:1516–1531.
15. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) *Proc Natl Acad Sci USA* 102:15418–15422.
16. Baker N, Holst M, Wang F (2000) *J Comput Chem* 21:1343–1352.
17. Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP (2005) *Proc Natl Acad Sci USA* 102:8933–8938.
18. Li H, Trotta CR, Abelson J (1998) *Science* 280:279–284.
19. Li H, Abelson J (2000) *J Mol Biol* 302:639–648.
20. Randau L, Calvin K, Hall M, Yuan J, Podar M, Li H, Soll D (2005) *Proc Natl Acad Sci USA* 102:17934–17939.
21. Calvin K, Hall MD, Xu F, Xue S, Li H (2005) *J Mol Biol* 353:952–960.
22. Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, Abelson JN (1997) *Cell* 89:849–858.
23. Trotta CR, Paushkin SV, Patel M, Li H, Peltz SW (2006) *Nature* 441: 375–377.
24. Song J, Markley JL (2007) *J Mol Biol* 366:155–164.
25. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graphics* 14:33–38.